MUTATIONS THAT AFFECT THE STRUCTURE AND INTERACTIONS OF THE CORE ANTIGEN OF HEPATITIS B VIRUS

FIONA JANE STEWART

Thesis submitted for the degree of Doctor of Philosophy

Institute of Cell and Molecular Biology

Faculty of Science

University of Edinburgh

1993

For Mum and Dad

Contents

Abstract

The nucleocapsid of Hepatitis B Virus is an icosahedral structure composed of 180 subunits of the viral Core Antigen (HBcAg). This protein is 183 amino acids in size and has a highly arginine-rich carboxy-terminal region. It can be expressed at high levels in *E. coli*, in which it forms nucleocapsid-like core particles which are morphologically indistinguishable from nucleocapsids isolated from infected individuals. The latter property was utilised in this work, the aim of which was to investigate the role of particular amino acids of HBcAg in the determination of the structure of the protein monomer and that of the core particle.

Previous work (Stahl *et al.*, 1982) had suggested the importance of the amino-terminal region of the protein in the determination of its structure. This was investigated further in this work by the use of site-directed mutagenesis to create a series of deletion and substitution mutations within this region. These were expressed in *E. coli* and the importance of amino acid 3 was demonstrated by immunological assay.

HBcAg contains four cysteines, all of which are completely conserved among mammalian hepadnaviruses. The role of these cysteines in HBcAg and core particle structure determination and stability was investigated by the use of site-directed mutagenesis to create a series of mutants in which cysteine codons were replaced by serine codons in several combinations, and in the context of both the full-length protein and of a truncated protein lacking the arginine-rich carboxy-terminal region. These proteins were produced in *E. coli* and formed particles indistinguishable from wild-type particles in structure, even when no cysteines were present. Their behaviour during non-reducing SDS-polyacrylamide gel electrophoresis indicated the presence of different disulphide bond complements in different mutants and allowed a model to be deduced for the arrangement of disulphide bonds within the core particle. In addition, the stability of the mutant core particles was found to depend on the number and nature of the cysteines present.

Finally, the role of the carboxy-terminal region of the core protein in terms of the determination of disulphide bond formation involving the remainder of the protein was investigated by the use of a series of proteins with the same amino-terminus but different carboxy-terminal regions. Non-reducing SDS-polyacrylamide gel electrophoresis of this series of proteins demonstrated a dependence of disulphide bond formation on the nature of the carboxy-terminal region of the protein.

Acknowledgements

I would like to thank Ken for his supervision, encouragement, enthusiasm and teaching since I came to Edinburgh. Also Heather and Sandra for all their help in the lab, both in the planning and execution of experiments, and to everyone in the lab for making it such a nice place to work.

I am also very grateful to Peter McCulloch for always responding immediately to frantic requests for large quantities of ^{125}I-labelled anti-HBc/HBe IgG; Peter Highton, Derek Notman and Stella Bury for electron microscopy; Margaret Daniel and Richard Ambler for protein work; Jean and Joan for washing mountains of glassware; Fiona Govan for rescuing me from Word Perfect disasters, David Melton for use of his group's computer and printer; Frank and Graham for excellent photography and a good laugh; and Heather, Sandra, Fiona Gray, Alice, Mike Dyson, Volker and Thomas for helpful discussions. I would also like to thank the Science and Engineering Research Council and the Molecular Biology Research Fund for funding.

Also, thanks to everyone who helped keep me going when I was ill, especially Ken and Noreen, Heather, Sandra, Fiona Gray, Patrick, Fiona, Shelagh, Jim, Nik, Mike Goman, Susan, Mary, Filippo, Chris, David Finnegan and John Scaife.

Finally, I will be eternally grateful to mum and dad for emotional and financial support in equal measure, and to Heather.

Abbreviations

Amp	ampicillin
Amp^r	ampicillin-resistant
ATP	adenosine triphosphate
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
bp	base pairs
BSA	bovine serum albumin
°C	degrees Celsius
CAH	Chronic Active Hepatitis
CAT	chloramphenicol acetyl transferase
CPH	Chronic Persistent Hepatitis
cpm	counts per minute
D	Daltons
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytosine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddTTP	dideoxythymidine triphosphate
DEPC	diethyl pyrocarbonate
dGTP	deoxyguanine triphosphate
DHBcAg	Duck Hepatitis B Virus core Antigen
DHBV	Duck Hepatitis B Virus
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	dinucleoside triphosphate
DR1	Direct Repeat 1
DR2	Direct Repeat 2

DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
E. coli	*Escherichia coli*
EDTA	diaminoethanetetra-acetic acid disodium salt
ER	endoplasmic reticulum
GSHV	Ground Squirrel Hepatitis Virus
HBcAg	Hepatitis B Virus core Antigen
HBeAg	Hepatitis B Virus e antigen
HBsAg	Hepatitis B Virus surface Antigen
HBxAg	Hepatitis B Virus x Antigen
HBV	Hepatitis B Virus
HHBV	Heron Hepatitis B Virus
HIV	Human Immunodeficiency Virus
IgG	immunoglobulin G
IgM	immunoglobulin M
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pairs
kD	kilodaltons
LTR	long terminal repeat
mRNA	messenger RNA
NBT	nitro blue tetrazolium
OD	optical density
OD_{650}	optical density at 650 nm
ORF	open reading frame
PCI	phenol/chloroform/isoamyl alcohol
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEG 6000	polyethylene glycol of molecular weight approximately 6000
pHCC	primary hepatocellular carcinoma
pHSA	polymerised human serum albumin
PMSF	phenylmethylsulphonamide

RNA	ribonucleic acid
RNase	ribonuclease
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SV40	Simian Virus 40
TEMED	N,N,N,N′-tetraethylenediamine
Tet	tetracycline
Tet^r	tetracycline-resistant
T_h	T helper cell
U	units
UV	ultra violet
V	volts
v/v	volume/volume
WHV	Woodchuck Hepatitis Virus
w/v	weight/volume
w/w	weight/weight

Amino Acid Abbreviations

	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

CHAPTER 1: Introduction

1.1 Hepatitis B Disease

Hepatitis B is a health problem of major global importance, with an estimated 5% of the world's population infected with the causative agent, Hepatitis B virus (HBV). The geographical distribution of HBV is uneven, with very low rates of infection in Europe (0.1-0.5%) and high rates (up to 15%) in parts of Africa and Asia. Methods of transmission vary correspondingly, in the West most often *via* injection of blood or blood-derived products, or by sexual contact, while in the East perinatal transmission is most common.

Hepatitis B virus is not itself cytolytic and the hepatitis resulting from infection is due to the host immune response to that infection, which varies from one individual to another, thereby resulting in the many different manifestations of hepatitis B. "Acute" hepatitis is the most common response and is characterised by diffuse inflammation of the liver. In most cases HBV antigens are difficult to detect in the liver, reflecting the active elimination of infected cells by the host immune response and subsequent resolution of the infection. However, in more severe acute cases, HBV antigens can be detected in a high percentage of hepatocytes. This indicates an inability to eliminate HBV-infected cells and in such cases there is a high propensity to progress to "chronic" hepatitis. This disease state occurs in 5-10% of infected individuals and is diagnosed when inflammation of the liver has persisted without improvement for 6 months. It can be divided into 2 further classes - "chronic active hepatitis" (CAH) and "chronic persistent hepatitis" (CPH). CAH is characterised by significant parenchymal damage and a higher number of hepatocytes containing Hepatitis B Core antigen (HBcAg) than Hepatitis B Surface antigen (HBsAg), while in CPH inflammatory infiltrates are present but there is minimal cell damage and the proportions of HBcAg and HBsAg are approximately reversed compared to CAH.

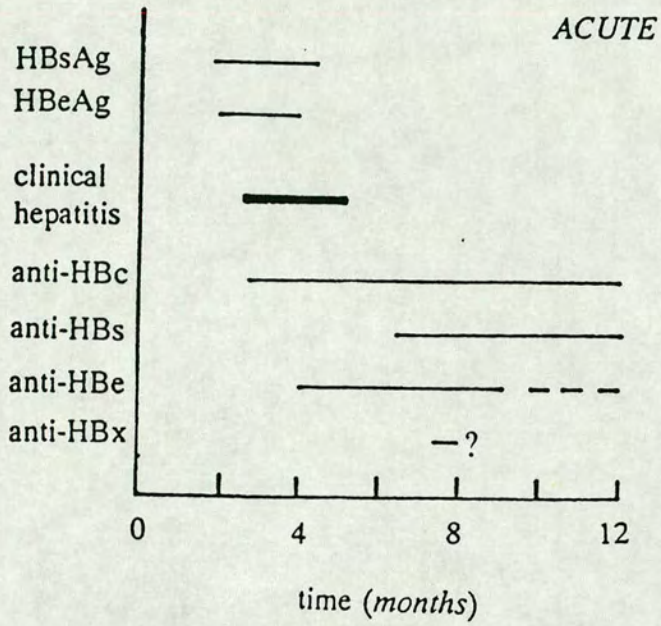
Various markers of infection can be detected in serum during HBV infection and this also reflects the difference in host immune response in the acute and chronic disease

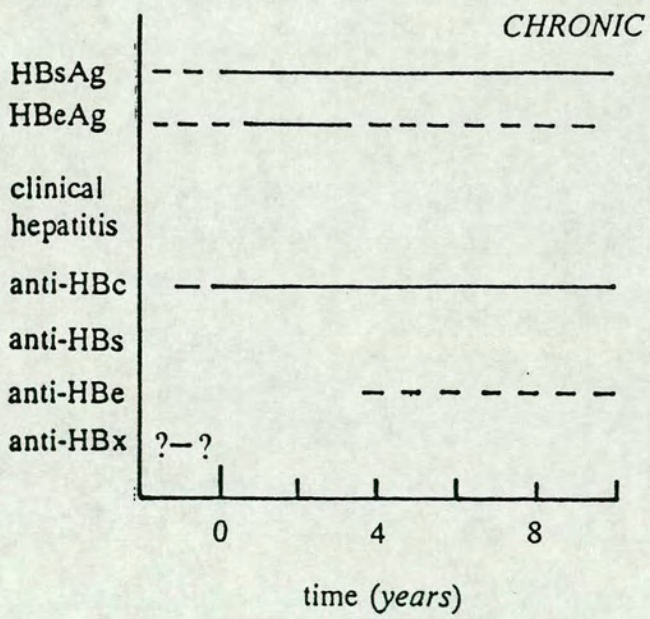
states, as shown in Figure 1.1. Detection of such markers in routine blood bank screening and general population screening has revealed a large number of chronically infected individuals without hepatitis B symptoms or signs of chronic liver disease who are thus termed "healthy" or "chronic carriers" and it has been estimated that as many as 50% of chronically-infected persons may be such healthy carriers. The host response to chronic liver injury results in cirrhosis of the liver which is associated with bands of fibrosis throughout the liver, which are, in turn, associated with nodules of parenchymal tissue. Necrosis of the hepatocytes stimulates liver regeneration, but lack of tight control of cell proliferation means that cell division continues beyond the level of replacement and, in the case of HBV infection, large hyperplastic nodules form. A combination of fibrosis and nodule formation can lead to both rupture of veins in the liver, causing massive haemorrhaging, and reduced access of oxygen and nutrients within the liver, resulting in hepatocyte dysfunction and subsequent liver failure. Distribution of HBV antigens within cirrhotic livers is heterogeneous, probably reflecting the different stages of maturation of individual nodules.

Finally, a strong association between HBV and primary hepatocellular carcinoma has been noted. Approximately 80% of primary hepatocellular carcinomas (pHCCs) in the world are associated with chronic HBV infection (W.H.O., 1983), and carriers of HBV are at least 100 times more likely to develop pHCC than non-carriers (Beasley *et al.*, 1981). In addition, pHCC has been demonstrated to develop in woodchucks, ground squirrels and ducks infected with the appropriate hepadnavirus, and hepadnavirus DNA has been found to be integrated into host DNA in a very large number of individuals with pHCC. However, such integration appears to be non-uniform and no region of the HBV genome or the host genome is consistently involved in integration events. The recent discoveries that the HBV-encoded proteins HBxAg and a truncated preS/S protein can act as transcriptional transactivators (Section 1.3.4) has led to the proposition that these proteins may be involved in tumorigenesis if expressed from integrated HBV DNA. In particular, it has been demonstrated recently that HBxAg acts by a mechanism involving an increase in the levels of protein kinase C in the cell (Kekule *et al.*, 1993). It is known that over-production of protein kinase C results in

Figure 1.1

The appearance of HBV antigens, and antibodies to them, in serum during acute and chronic infection. The variable nature of the appearance of these markers is indicated by the broken lines.


ACUTE
HBsAg
HBeAg
clinical hepatitis
anti-HBc
anti-HBs
anti-HBe
anti-HBx
—?
0
4
8
12
time (months)


CHRONIC
HBsAg
HBeAg
clinical hepatitis
anti-HBc
anti-HBs
anti-HBe
anti-HBx
?—?
0
4
8
time (years)

disordered growth of mammalian cells and so the predicted oncogenicity of HBxAg may be mediated in this manner. A further important factor common to chronically-infected individuals is damage to the liver, which stimulates regeneration of liver cells. Such a rapidly-growing population of liver cells will provide greater opportunities for alterations to cells which may lead to tumorigenesis.

1.2 Virion and Genome Structure

The initial report of the association of the so-called "Australia" antigen with Hepatitis B (Blumberg *et al.*, 1965) was followed by the observation under the electron microscope that this antigen was present in serum in the form of 20-25nm virus-like particles (Bayer *et al.*, 1968; Gerin *et al.*, 1969). However, it was subsequently discovered (Dane *et al.*, 1970) that such particles, while by far the most abundant of virus-derived particles present during infection, are not the actual virions. The latter are 42nm in diameter and have a "double-shelled" appearance under the electron microscope. That this 42nm particle is in fact the infectious agent of Hepatitis B was shown by Robinson and Luttwick (1976).

Hepatitis B virus is now regarded as the prototype member of the hepadnavirus (for "hepatic" and "DNA") family which also includes viruses isolated from the Eastern Woodchuck (Summers *et al.*, 1978), the Beechey Ground Squirrel (Marion *et al.*, 1980), the Pekin Duck (Mason *et al.*, 1980) and most recently the Grey Heron (Sprengel *et al.*, 1988). These viruses have been named, respectively, Woodchuck Hepatitis Virus (WHV), Ground Squirrel Hepatitis Virus (GSHV), Duck Hepatitis B Virus (DHBV) and Heron Hepatitis B Virus (HHBV), and are grouped together on the basis of their hepatotropism, similar virion morphology and common genome size, structure, and replication strategy. Hepadnaviruses also show a very narrow host range and the human virus, for example, infects only humans and a few higher primates, of which the chimpanzee is the best known example.

The structure of the HBV virion is represented in Figure 1.2A. The envelope consists of host-derived lipid interspersed with HBsAg molecules. This envelope surrounds the virus nucleocapsid, an icosahedral structure of 27nm diameter composed of subunits of HBcAg, and inside this are contained the viral polymerase, the viral genome and the genome-linked protein. Titres of Dane particles vary between infected individuals from 10^3 - 10^6 particles/ml. However, far in excess of this are the subviral HBsAg particles viewed initially by Bayer *et al.* (1968) and Gerin *et al.* (1969). These are essentially "empty envelopes", consisting only of lipid and HBsAg, but with a HBsAg composition different from that of Dane particle envelopes (Section 1.4.1), and are found in 2 forms - 22nm diameter spheres and filaments of the same diameter and variable length. (Figure 1.2b). The biological function of these non-infectious particles is unclear, but they may act as "decoys" by attracting neutralising antibodies. Indeed, acutely-infected patients often have arthritis due to to deposition of subviral particle/antibody complexes in their joints (Gocke, 1975).

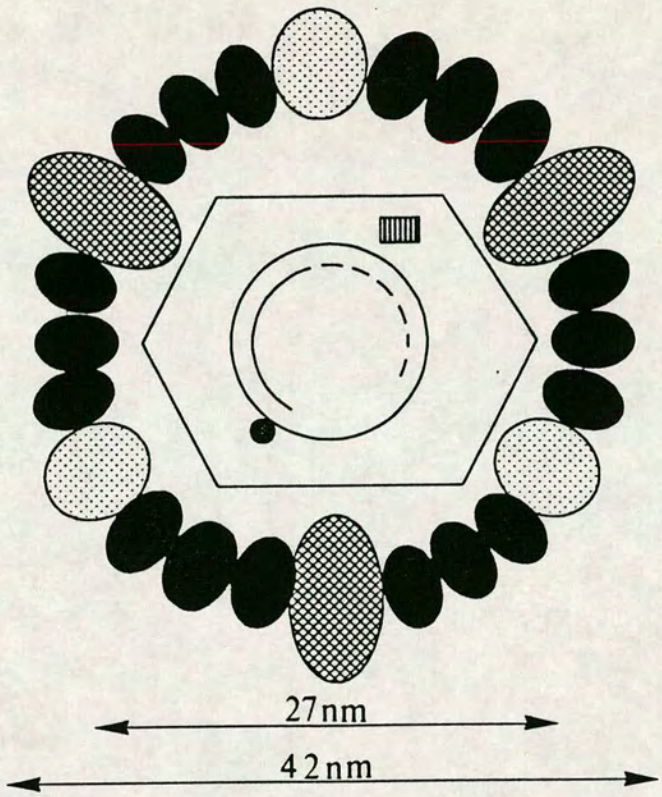
The circular nature of the genome of HBV was determined by electron microscopic examination by Robinson *et al.* (1974) and Overby *et al.* (1975) and its characteristic and unusual partially double-stranded nature was deduced by Summers *et al.* (1975) and is shown in Figure 1.3. The long strand serves as the template for viral transcription and is thus termed the minus strand. In the HBV isolate used in this work it is 3190 bases in length (Burrell *et al.*, 1979; Pasek *et al.*, 1979; Pugh *et al.*, 1986). The single-stranded gap varies from 20-80% of unit length and the 5′ end of the short strand is fixed while the 3′ end is variable. The long strand is not covalently-closed and is terminally redundant. The circular nature of the genome is maintained as the 5′ end of the short strand is approximately 300 nucleotides downstream of the 5′ end of the long strand. Annealing of the long and short strands creates a tail from the 8bp overlap to which a protein is covalently attached (Gerlich and Robinson, 1980). The 5′ end of the short strand is also unusual in its covalent attachment to a short capped oligoribonucleotide, first demonstrated for DHBV (Lien *et al.*, 1986), and determined to be approximately 17 bases in length.

Figure 1.2

A. Structure of the HBV virion.
The viral envelope consists of surface antigen (S protein ● , M protein ◌ , and L protein ▮) embedded in a lipid bilayer. This surrounds the nucleocapsid (⬡), composed of Core antigen, which in turn surrounds the viral genome, polymerase ▤ and genome-linked protein ● .

B. Structure of subviral spheres and filaments. These are similar to empty viral envelopes, but have a smaller diameter and different surface antigen composition. Filaments have the same diameter as spheres, and are of variable length.

A



B

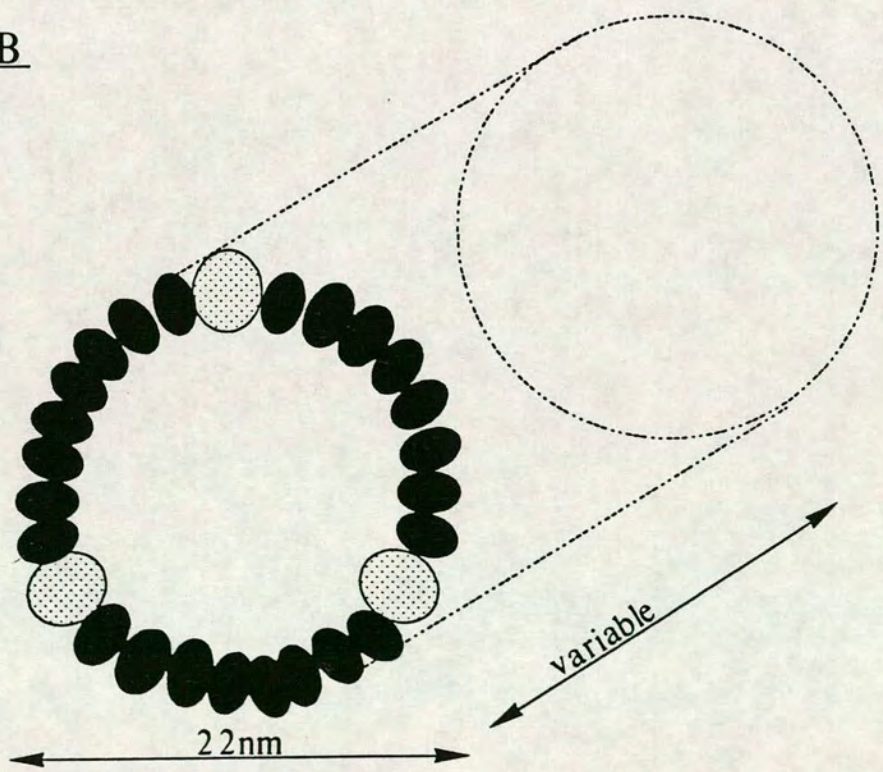
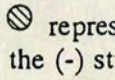
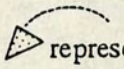


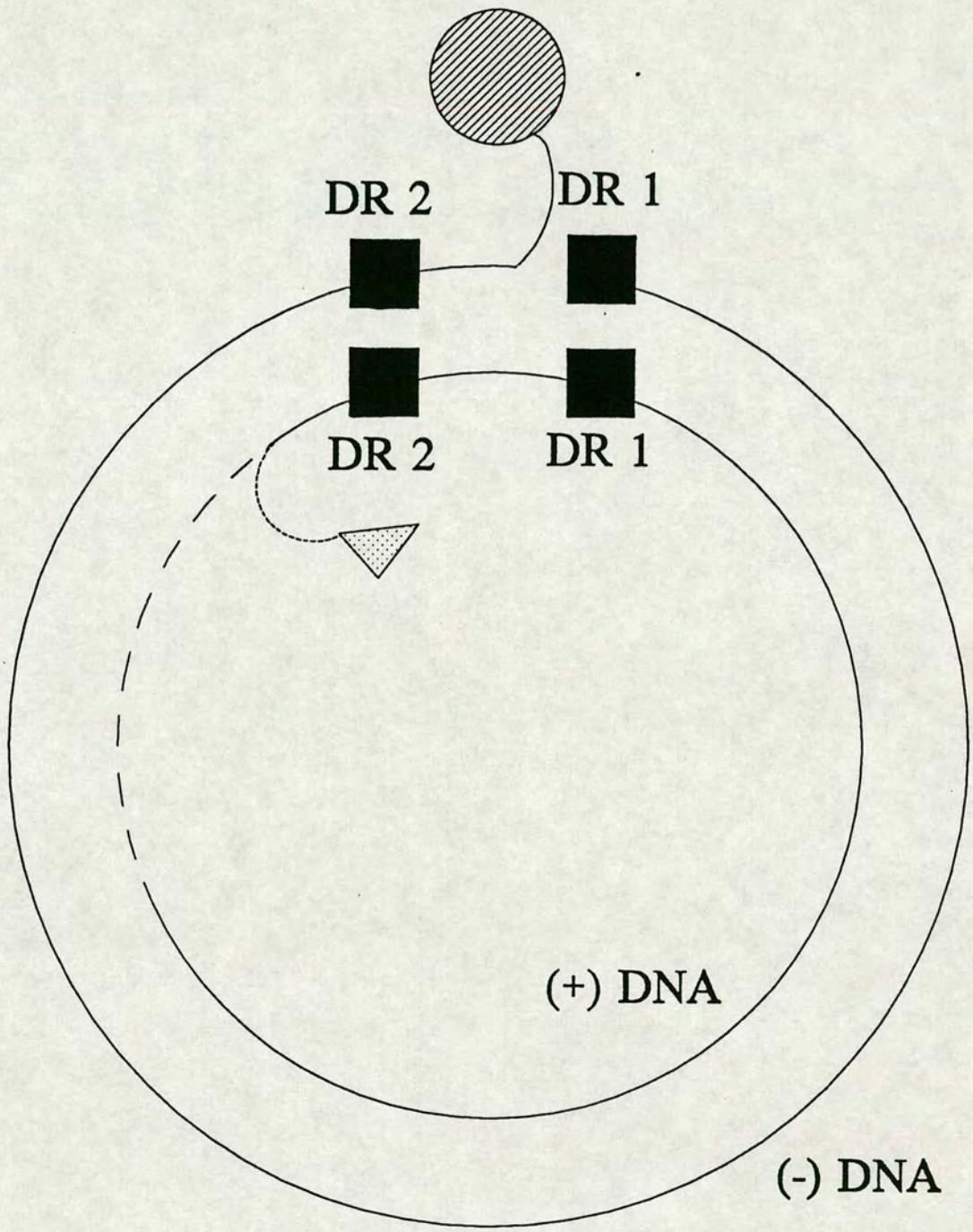
Figure 1.3

Structure of the partially double-stranded DNA genome of HBV.

 represents the genome-linked protein attached to the terminus of the redundant portion of the (-) strand.

 represents the capped oligoribonucleotide attached to the 5′ terminus of the (+) strand.

DR = Direct Repeat.


DR 2
DR 1
DR 2
DR 1
(+) DNA
(-) DNA

Within the genome are two 11bp directly repeated sequences, termed DR1 and DR2 which are located close to the 3′ end of the minus strand and the 5′ end of the plus strand, respectively. These direct repeats, together with the genome-linked protein and the covalently-attached oligoribonucleotide are important for replication of the viral genome and will be discussed in section 1.5.

1.3 Genome Organisation and Viral Transcription

At 3.2kb the genome of HBV is one of the smallest DNA virus genomes known. Indeed, it was initially suggested (Overby *et al.*, 1975) that a helper virus may be required to programme the life cycle of the virus. However, the cloning in *E. coli*, and subsequent sequencing, of the genomes of several HBV isolates (Galibert *et al.*, 1979; Pasek *et al.*, 1979; Valenzuela *et al.*, 1980; Fujiyama *et al.*, 1983; Ono *et al.*, 1983; Kobayashi and Koike, 1984; Okamoto *et al.*, 1986) has revealed a very full use of the entire genome, with extensively overlapping coding sequences, thus allowing infection to proceed without the aid of a helper virus.

The HBV genome was found to contain four major open reading frames (ORFs) in the minus strand, termed S, C, P and X, utilising all three translational frames. No conserved open reading frames of significant length are present in the plus strand. However, as there are at least seven viral gene products, the transcriptional strategy is complex, with several transcriptional control regions, which are also themselves transcribed during active infection, and necessarily within a short distance of each other. The organisation of the HBV genome is shown in Figure 1.4, which demonstrates the multiple capacity of the majority of nucleotide sequences in the genome. In this thesis, the numbering system used for HBV DNA is that of Pasek *et al.* (1979), whereby nucleotide number 1 is the first nucleotide of the coding sequence of HBcAg.

Figure 1.4

HBV genes and transcripts.

The genome is represented in a linear manner and nucleotide positions of the beginning and end of each open reading frame are shown.

spI = preS1 promoter

spII = pres2/s promoter

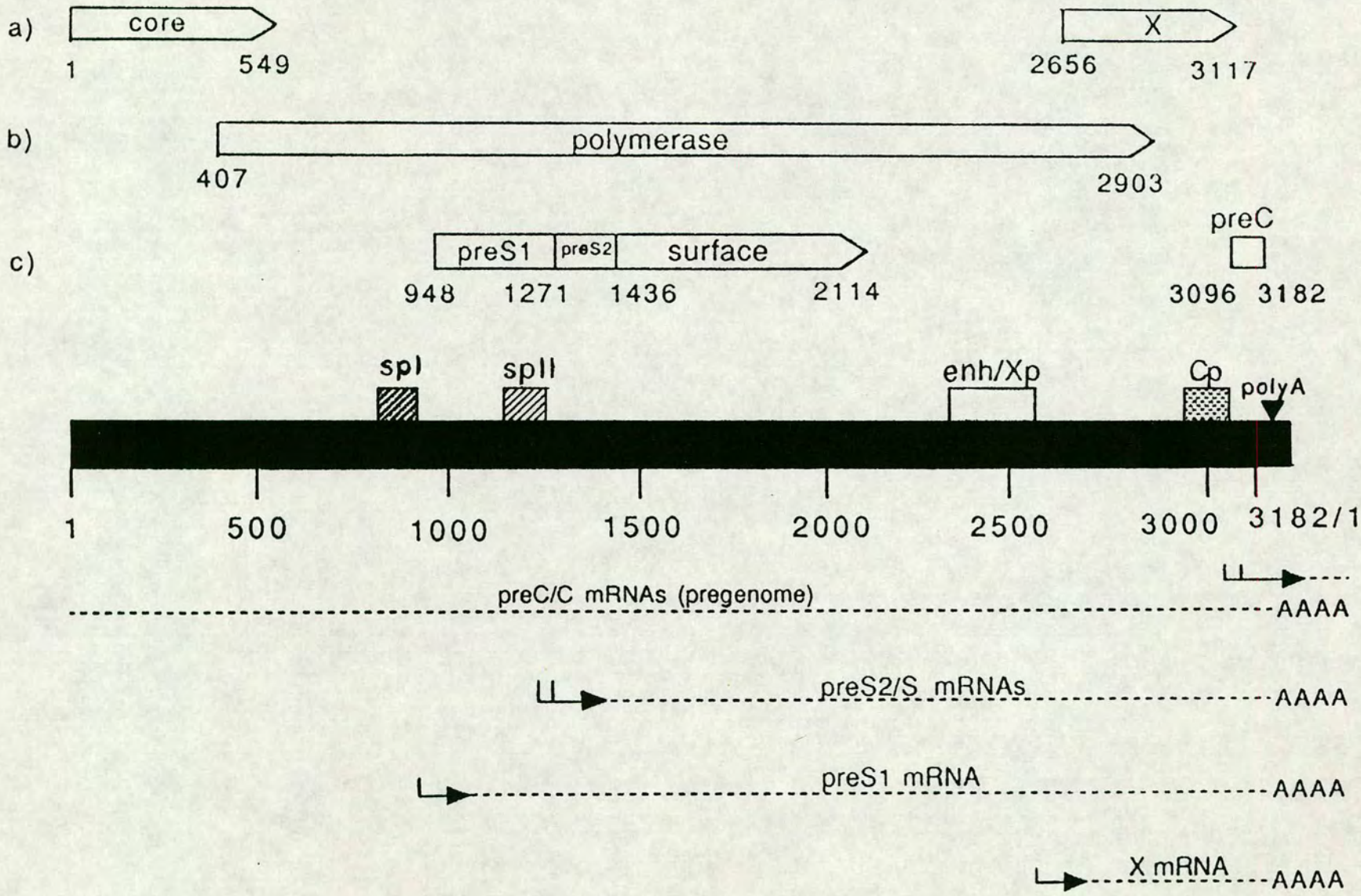
enh/Xp = enhancer/X promoter

Cp = core promoter

▾ = HBV polyadenylation signal at nucleotide position 16.

Viral transcripts are represented by dashed lines and the transcription initiation sites by vertical lines.

(From Rossner, 1991)


a)
core
1
549
X
2656
3117
b)
polymerase
407
2903
c)
preS1
preS2
surface
preC
948
1271
1436
2114
3096
3182
spI
spII
enh/Xp
Cp
polyA
1
500
1000
1500
2000
2500
3000
3182/1
preC/C mRNAs (pregenome)
AAAA
preS2/S mRNAs
AAAA
preS1 mRNA
AAAA
X mRNA
AAAA

Initial studies of transcription of HBV genes were hampered by the lack of a cell culture system for propagation of HBV. Therefore the earlier studies involved examination of transcripts from a cell line derived from a human hepatoma (Chakraborty *et al.*, 1980; Edman *et al.*, 1980), rodent cells transformed with multiple copies of the HBV genome (Pourcel *et al.*, 1982; Gough and Murray, 1982) and the livers of chimpanzees infected with HBV (Cattaneo *et al.*, 1983; 1984). More recently, however, cell lines derived from human hepatomas that do not have integrated HBV DNA, such as HepG2 (Aden *et al.*, 1979) and HuH7 (Nakabayashi *et al.*, 1982) have been found, after transfection with HBV DNA, to support HBV replication (Sureau *et al.*, 1986; Sells *et al.*, 1987; Yaginuma *et al.*, 1987) and to produce infectious virions (Acs *et al.*, 1987; Sureau *et al.*, 1988). Primary duck hepatocytes have also been cultured and shown to be permissive for DHBV infection (Tuttleman *et al.*, 1986b). However, the cells remain in a differentiated state for only a short time, thus limiting the experiments that can be performed.

The HBV genome contains only one polyadenylation signal and therefore all messenger RNAs (mRNAs) terminate at the same position. The RNA species produced by HBV can be divided into two classes - genomic and sub-genomic transcripts. As the genome replicates *via* reverse transcription of an RNA intermediate (Section 1.5), genomic RNAs serve the dual role of template for (i) translation to produce viral polypeptides, and (ii) reverse transcription to produce genomic DNA. In contrast, sub-genomic transcripts act only as templates for translation.

1.3.1 Genomic Transcripts

These RNAs are approximately 3.3kb in length and are therefore terminally redundant. Their synthesis is under control of the C (or genomic) promoter which directs transcription of two subsets of genomic transcripts (Figure 1.4). The larger of these initiates upstream of the pre-core start site and therefore encodes the precore/core polypeptide precursor of HBeAg. The more abundant transcript initiates at nucleotide

position 3001+/-2 (Will *et al.*, 1987), immediately downstream of the precore start site, and therefore cannot encode the pre-core region. This mRNA encodes the core and polymerase proteins of the virus (Ou *et al.*, 1990). As these genes overlap by 142bp it was initially suggested that the polymerase may be produced by processing of a core/pol fusion protein resulting from ribosomal frameshifting during translation of this mRNA, in a manner analagous to retroviruses. However, it has subsequently been shown that translation of the polymerase sequences in DHBV begins at the first internal AUG in the polymerase gene (Chang *et al.*, 1989b; Schlicht *et al.*, 1989b; Kawamoto *et al.*, 1990) and this has also been demonstrated to occur in HBV (Jean-Jean *et al.*, 1989c; Kawamoto *et al.*, 1990; Roychoudhury and Shih, 1990). In addition to its protein coding capacity, this mRNA is also indispensable for viral replication since it is the template for reverse transcription to produce the viral DNA genome (Will *et al.*, 1987). This will be discussed in further detail in section 1.5.

1.3.2 Sub-genomic Transcripts

Immediately upstream of the S gene are two regions termed preS1 and preS2 (Figure 1.4). Translation of these regions results in three surface proteins: S, preS2/S and preS1/preS2/S. These are not produced by post-translational processing of the latter, but by differential transcription of the preS and S ORFs. There are two classes of surface mRNAs - those of approximately 2.4kb in length, and those of approximately 2.1kb. The former initiate approximately 45bp upstream of the first initiation codon of the preS1 ORF and therefore include coding capacity for preS1/preS2/S (Laub *et al.*, 1983; Rall *et al.*, 1983; Ou and Rutter, 1985). However, this is a minor surface mRNA class, comprising only an estimated 2% of total surface mRNAs (Ou and Rutter, 1985).

The more abundant transcripts are approximately 2.1kb in length. These have heterogeneous initiation sites within the preS region (Cattaneo *et al.*, 1983; 1984; Laub *et al.*, 1983; Yaginuma *et al.*, 1987), clustered around the AUG codon of preS2 (Ou

and Rutter, 1985). Transcripts initiating upstream of the preS2 AUG can encode the preS2/S protein, while those initiating downstream encode S only. These transcripts are produced in approximately equal amounts (Ou and Rutter, 1985; Yaginuma *et al.*, 1987), but the predominant form of surface antigen produced in infected hepatocytes is S, suggesting inefficient translation initiation at the preS2 AUG. This has been found to be the case when this mRNA was injected into *Xenopus* oocytes (Standring *et al.*, 1986) or expressed in mammalian cells under control of a heterologous promoter (McLachlan *et al.*, 1987; Ou and Rutter, 1987; Molnar-Kimber *et al.*, 1988).

An additional set of transcripts of approximately 1kb in length has been mapped to the X gene of HBV, both in mammalian cells transfected with the complete HBV genome (Gough, 1983; Zelent *et al.*, 1987; Koike *et al.*, 1989), or with a fragment of the HBV genome containing the surface and X ORFs (Simonsen and Levinson, 1983; Saito *et al.*, 1986; Siddiqui *et al.*, 1986; Bulla and Siddiqui, 1988, 1989). The X polypeptide is proposed to be produced *in vivo* by translation of this 1kb mRNA and not by initiation of translation at an internal site in any other HBV mRNA. There is one report of an X-specific RNA detected *in vivo*, but the RNA in this case was 0.65kb and postulated to be the X mRNA lacking its polyA tail (Kaneko and Miller, 1988).

All of the above mRNAs are non-spliced and for many years it was thought that this was true for all HBV transcripts. Recently, however, there have been several reports of a singly-spliced transcript of approximately 2.2kb (Su *et al.*, 1989; Suzuki *et al.*, 1989; Terre *et al.*, 1991; Wu *et al.*, 1991) and a doubly-spliced transcript of approximately 2.6kb (Chen *et al.*, 1989; Suzuki *et al.*, 1990; Terre *et al.*, 1991; Wu *et al.*, 1991), which share the 5′ and 3′ extremities of the genomic transcripts and have been detected in HBV DNA-transfected HepG2 cells and in HBV-infected liver. Such spliced transcripts are not essential for viral replication but it has been reported (Terre *et al.*, 1991) that these transcripts are packaged into viral nucleocapsids and reverse transcribed *in vivo*, contributing to the population of defective viruses often found in

chronically-infected individuals. However, the importance and biological function of these transcripts remains unclear.

1.3.3 Regulation of Transcription

Cis-acting elements

Four promoter elements have been functionally determined to be present within the HBV genome. Functional core/genomic promoter elements have been found to be contained in the region 2975-3100 and no TATA or CAAT box homology is found within this region. (Honigwachs *et al.*, 1989; Yaginuma and Koike, 1989). However, approximately 25 nucleotides upstream of the initiation site of preS1/preS2/S mRNA is a TATA-like sequence: 876-TATATAA-882 (Rall *et al.*, 1983) which corresponds to the position of such an element in most eukaryotic promoters. The minimal preS1/preS2/S promoter element was determined to contain this element and to lie between positions -90 and +35, relative to the transcription initiation site. (Raney *et al.*, 1990).

As in the case of the core/genomic promoter, sequences upstream of the start sites of the preS2/S mRNAs have no TATA or CAAT-like sequences. However, sequences upstream of these start sites do have homology to another eukaryotic TATA-less promoter - the SV40 late promoter (Brady *et al.*, 1982), at a similar position relative to initiation of transcription (Cattaneo *et al.*, 1983). Subsequent reports (Raney *et al.*, 1989; DeMedina *et al.*, 1988) suggest, however, that this region of homology is not essential for promoter activity and that upstream and downstream sequences may also be important. The fourth promoter element of the HBV genome lies upstream of the X gene and directs transcription of its 1kb mRNA. It has been identified to lie within 200 nucleotides 5′ to the X ORF and can direct transcription of a heterologous reporter gene (Siddiqui *et al.*, 1987; Treinin and Laub, 1987; Wollersheim *et al.*, 1988). This region also contains the viral enhancer, removal of which drastically reduces X promoter activity. The viral enhancer can therefore be thought of as an

integral part of the X promoter.

Two enhancer elements have been identified within the HBV genome - the first is that described above, within the X promoter region, and is termed Enhancer I (Shaul *et al.*, 1985). Transcription from all HBV promoters has been shown to be increased in the presence of this enhancer element and its deletion from HBV genomic DNA resulted in reduced initiation of transcription from all HBV promoters in transiently transfected cells (Hu and Siddiqui, 1991). A second region, termed Enhancer II, was described by Yee (1989) as encompassing nucleotides 2966-3056. This 88bp fragment, located upstream of the C gene, has subsequently been shown to enhance transcription from the core, preS and S promoters of HBV (Su and Yee, 1992).

Negative regulation by *cis* elements within the HBV genome has also been observed. The preS1 promoter was shown to be negatively regulated by DNA sequences containing the downstream preS2 promoter region (Bulla and Siddiqui, 1989). This repression was demonstrated to be at the level of transcription and not to be effected by preS2/S or S proteins. This may also contribute to the relative paucity of preS1 transcripts.

Finally, all hepadnaviral RNAs terminate at the same position, approximately 20 nucleotides downstream of the conserved sequence 16-TATAAA-21 (Cattaneo *et al.*, 1983; 1984; Simonsen and Levinson, 1983), which is very similar to the polyadenylation signal AATAAA found in eukaryotes (Proudfoot and Brownlee, 1976). However, this sequence alone has been shown to be insufficient for efficient mRNA processing and that approximately 30 nucleotides of downstream sequence are also required (Simonsen and Levinson, 1983). In order to produce the greater-than-genome-length transcripts required for genome replication and precore/core, HBcAg and viral polymerase production; the polyadenylation signal must be ignored on first passage, but it must be recognised and used on the second encounter. This is thought to be effected by the one nucleotide difference between the HBV polyadenylation signal and the consensus signal, and by the participation of multiple upstream

elements. As only one of the major upstream elements is present between the initiation site of pregenomic RNA and the processing site, this was thought to explain why the processing signal is used so inefficiently (Russnak and Ganem, 1990; Russnak, 1991). However, a recent report (Cherrington *et al.*, 1992) demonstrates that this element alone is sufficient for polyadenylation and that it is rather the short distance from the cap site of pregenomic mRNA to the UAUAAA sequence that causes bypass of the 5′ poly(A) site, as this supression can be relieved by increasing the distance to 230-400 nucleotides.

1.3.4 Transcription regulation by viral gene products

The X gene product of HBV has recently been shown to act as a transcriptional transactivator of many viral and cellular promoters. Most importantly for the viral life cycle, HBxAg has been shown to regulate transcription from the HBV core promoter/enhancers, the HBV preS1 promoter/enhancers and the X promoter/enhancer. (For a recent comprehensive review see Rossner, 1992). In addition, human hepatoma-derived cells tranfected with an HBV genome mutant in HBxAg had reduced levels of viral transcription (Colgrove *et al.*, 1989; Koike *et al.*, 1989). Therefore the X antigen is proposed to be involved in HBV transcriptional regulation *in vivo*. A second class of HBV-derived transcriptional transactivator has recently been identified. (Kekule *et al.*, 1990; Caselmann *et al.*, 1990). 3′ truncated preS2/S sequences were found to stimulate in *trans* the SV40 promoter attached to the CAT reporter gene. Such transactivation was not seen with intact preS2/S sequences. As these sequences were initially identified as being the integrated HBV DNA sequences in a human hepatoma-derived cell line, it has been proposed that this truncated protein may be involved in the development of primary hepatocellular carcinoma. However, as such a truncated protein has not been detected during normal HBV infection, it is unlikely to have a role in the transcriptional transactivation of HBV promoters in the normal life cycle of the virus.

Finally, negative regulation of a cellular gene (human β-interferon) by HBV has been

reported (Twu and Schloemer, 1989) and it has been proposed that this function is performed by the C gene product, HBcAg (Whitten *et al.*, 1991). A regulatory function for this protein is also suggested by the presence in the HBcAg sequence of a nuclear targetting signal (Ou *et al.*, 1989). However, this observation has been limited to the β-interferon gene and there is no evidence that such negative regulation is relevant to the HBV life cycle, although it may affect virus/cell interactions.

1.3.5 Tissue-specificity

While hepadnaviruses are largely liver-specific with regard to infection, non-hepatocytes have been found to support hepadnavirus replication both in cultured cells (Korba *et al.*, 1988; Galun *et al.*, 1992), infected individuals (Blum *et al.*, 1983; Korba *et al.*, 1989; Ogston *et al.*, 1989) and HBV-transgenic mice (Perfumo *et al.*, 1992). It has also been noted that the pattern of extra-hepatic infection in chronically- and acutely-infected woodchucks is different (Korba *et al.*, 1989) but it is not known whether this is a cause or a symptom of the different pathologies. The mechanism of this relative hepatotropism is unclear, but association between viral transcriptional control elements and liver-specific factors is thought to be likely and several candidate elements and factors have been identified.

While the preS2/S promoter is not stringently liver-specific and can be utilised in mouse L cells (Pourcel *et al.*, 1982), its expression has been found to be higher in the liver and kidney of HBV transgenic mice (DeLoia *et al.*, 1989). In contrast, the preS1 promoter is highly liver-specific, and has been found to express 5-90 times more efficiently in differentiated hepatoma-derived liver cell lines than in either de-differentiated liver cells or non-liver cells (Raney *et al.*, 1990). In addition, it has been shown that a liver-specific factor (HNF-1) binds to a HNF-1 binding site within the preS1 promoter (Chang *et al.*, 1989a) and that this factor is also responsible for the differentiation-specific expression from this promoter. The core/genomic promoter is also expressed to a greater extent in differentiated hepatoma-derived cells (Raney *et al.*, 1990) and this partial tissue-specificity has been shown to become almost

complete when the core promoter and enhancer I are present together (Honigwachs *et al.*, 1989). A specific factor found in hepatoma-derived cells has been demonstrated to interact with an essential element within the core promoter (Yaginuma and Koike, 1989; Karpen *et al.*, 1988) while such a factor is not present in the non-hepatic HeLa or NIH 3T3 cells. However, negative regulation by factors present in non-liver cells may also be involved. Guo *et al.* (1993) found that the transcriptionn factor HNF-4, which is enriched in hepatocytes, stimulated transcription of a core promoter containing a HNF-4 binding site by approximately 20-fold but that the presence of sequences upstream of the HNF-4 binding site in the core promoter prevented stimulation of transcription by HNF-4 in HeLa cells, but not in hepatoma-derived HuH7 cells.

The enhancer elements of HBV have been studied extensively with regard to tissue-specificity and while Vannice and Levinson (1988) found no cell-type specificity of enhancer I this has been disputed by several other reports. Although enhancer I is functional in non-liver cells, expression is dramatically increased in liver-derived cells (Karpen *et al.*, 1988, Antonucci and Rutter, 1989; Honigwachs *et al.*, 1989; Huan *et al.*, 1992). Protein-binding sites within this enhancer were identified by Karpen *et al.* (1988) and it has recently been shown that within the enhancer I is a retinoid X response element, to which RxR binds (Huan *et al.*, 1992). This factor is highly enriched in liver but is still present in other cell types, possibly providing an explanation for the "leaky" cell type-specificity of this element. In contrast, enhancer II is stringently liver-specific and possibly differentiation stage-specific (Yee, 1989; Su and Yee, 1992). These specificities have recently been confirmed by Tay *et al.* (1992), who have also identified an apparently liver-specific protein (E2BP), which can transactivate enhancer II both in liver-derived cell lines, and in non-hepatic cell lines when both enhancer II and E2BP are transiently transfected. However, maximal stimulation is not seen with E2BP alone and two as yet unknown proteins found to bind to enhancer II in addition to E2BP may be required for optimal enhancer effect.

1.4 Viral Polypeptides

1.4.1 Surface Antigens

As described above, the surface antigen of HBV was originally termed the Australia antigen and was first found in the context of subviral particles, the overproduction of which is a hallmark of HBV infection. It was subsequently also found to be present on the surface of HBV virions (Dane *et al.*, 1970) and therefore was termed "surface antigen".

Examination of the S ORF reveals three in-frame translation initiation codons, all of which are utilised during natural infection to produce three polypeptides. The largest of these is termed the L protein, which initiates at the first AUG and thus consists of regions preS1, preS2 and S (Heerman *et al.*, 1984). It exists in both unglycosylated (p39) and glycosylated (gp42) forms and is also acylated at its amino terminal glycine residue with myristic acid (Persing *et al.*, 1987). It is the least abundant of the three forms of surface antigen and preS1-containing proteins represent only 10-20% of the surface proteins in HBV virions and subviral filaments, and only 1-2% of the surface proteins in the 20nm diameter spherical subviral particles.

The middle surface antigen, the M protein, initiates at the second AUG and therefore consists of the preS2 and S regions. It also exists in two forms. Gp33 is glycosylated only at amino acid 4 of the preS2 domain while gp36 is additionally glycosylated within the S domain (Stibbe and Gerlich, 1983). The proportion of preS2/S proteins is similar in all three particulate forms and is estimated to be approximately 10% (Heermann *et al.*, 1984).

The smallest surface antigen, the S protein, initiates at the third AUG and therefore contains no preS sequences. It is the most abundant form of HBsAg in the circulation of infected individuals and is present in both un-glycosylated (p24) and glycosylated (gp27) forms. The S protein is estimated to constitute approximately 70% of the

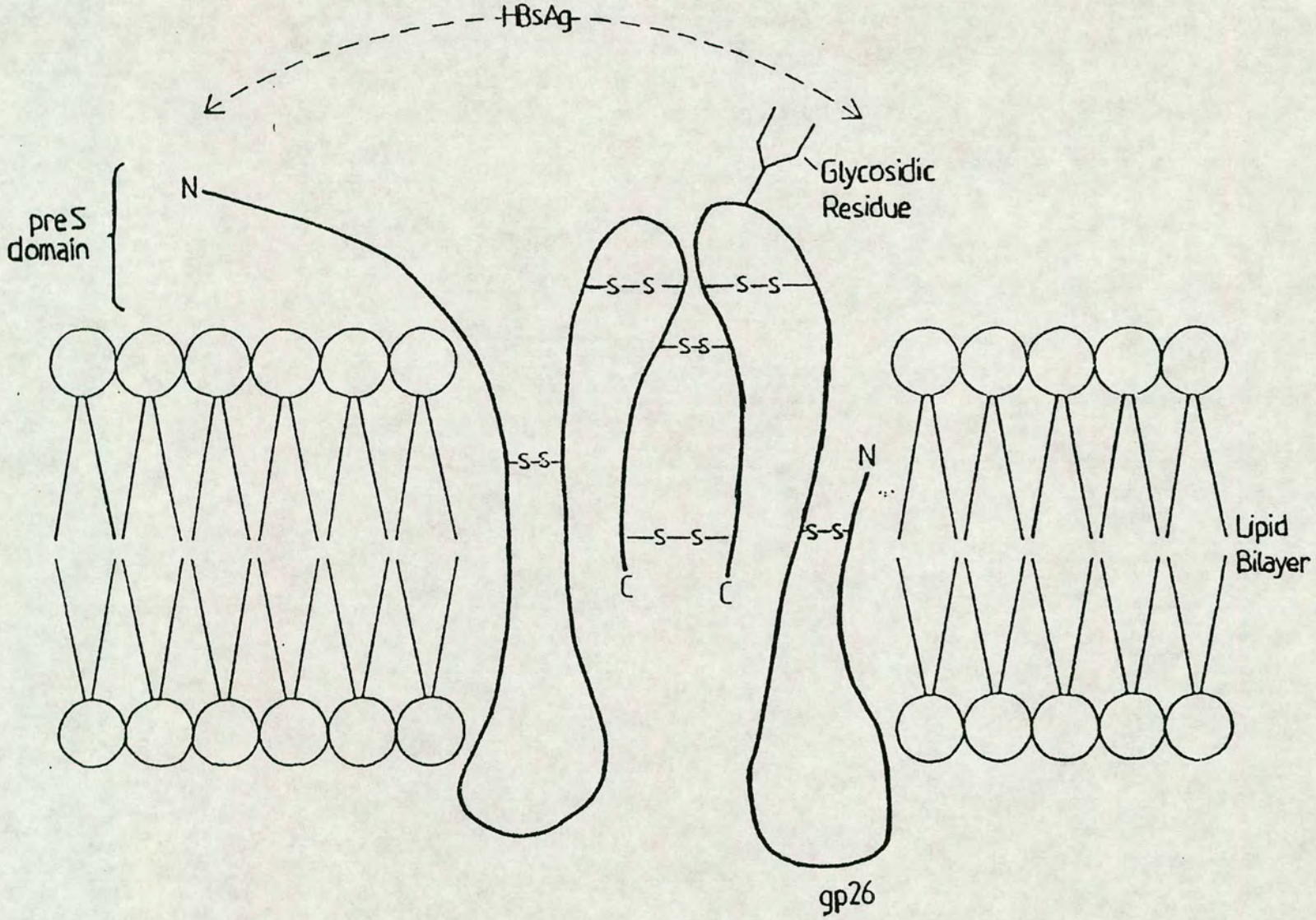
HBsAg present in Dane particles and subviral filaments (Ganem, 1991), and an even higher proportion in 20nm subviral spheres (Heermann *et al.*, 1984). Indeed, cells expressing only the S protein have been found to be capable of producing 20nm particles of apparently normal morphology (Liu *et al.*, 1982; Laub *et al.*, 1983), thus indicating that all the necessary information for 20nm particle formation lies within the S domain. This is not the case for virion formation, however, for which both the L protein and the S protein appear to be necessary (Bruss and Ganem, 1991; Ueda *et al.*, 1991). The role of the M protein in virion assembly is currently unclear. Ueda *et al.* (1991) detected no Dane particle formation in the absence of the M protein, while Bruss and Ganem (1991) found the M protein to be unnecessary for virion formation, and, in addition, replacement of the majority of the preS2 domain with foreign sequence did not prevent incorporation into the virion envelope. The non-essential nature of the M protein is also suggested by reports of the frequent occurrence in chronically-infected patients of a dominant virus population whose genomes cannot encode preS2-containing proteins (Fernholz *et al.*, 1991; Santantonio *et al.*, 1992).

The high proportion of cysteine residues within the HBsAg sequence (14 per 226 amino acid monomer) suggests the possibility of both intra- and inter-molecular disulphide bonds. The presence of inter-molecular disulphide bonds was confirmed by Mishiro *et al.* (1980) who demonstrated that the spherical subviral particles contain a polypeptide of approximately 49kD. This polypeptide exhibited all antigenic determinants of subviral particles, and in the presence of β-mercaptoethanol was cleaved into two polypeptides of 22kD and 27kD respectively. This would suggest that the predominant structural unit of surface antigen consists of a dimer of the S protein, one glycosylated and the other unglycosylated, in agreement with the observation that the two forms of this protein are present in approximately equal amounts in both virions and subviral particles (Heermann *et al.*, 1984). HBsAg has distinct hydrophobic and hydrophilic regions which have led to the proposition of a model for its structure within the HBV envelope (Figure 1.5), which is supported by experimental mapping of exposed sites (Peterson *et al.*, 1982). This structure accommodates dimer formation and also the preS regions, both of which have been

Figure 1.5

Proposed structure of HBsAg in the outer envelope of the HBV virion.

(From Ashton-Rickardt, 1988)


HBsAg
Glycosidic Residue
pre S domain
N
S-S
S-S
S-S
S-S
S-S
S-S
N
C
C
Lipid Bilayer
gp26

shown experimentally to be exposed in Dane particles and spherical particles (Stibbe and Gerlich, 1983; Heermann *et al.*, 1984; 1987).

As HBsAg is the viral antigen residing on the surface of the virus particle, it is the protein to which neutralising antibodies would be expected to be directed. This is indeed the case and the original HBV vaccine consisted of subviral particles purified from plasma of chronic carriers of the virus (Hilleman *et al.*, 1978). The only HBV proteins present in this vaccine are the surface proteins and it was found to be safe and effective as a vaccine (Szumess *et al.*, 1980). However, its high cost and the perceived danger of using material, although highly purified, from an infected individual led to the development of an alternative vaccine. This was made possible by the ability of yeast (*Saccharomyces cerevisiae*) after expression of a foreign S gene to produce subviral particles of apparently normal morphology, which could be produced and purified on a large scale, reliably free from contamination. Such a preparation was shown to protect chimpanzees upon subsequent challenge with HBV (Murray *et al.*, 1984) and, following clinical trials, is now in general use for vaccination of human populations.

Characterisation of antibodies to HBsAg defined distinct epitopes within HBsAg (Hirschman *et al.*, 1969). These were further defined to be (i) the *a* determinant, which is common to all HBV isolates and (ii) two pairs of sub-type determinants - *d/y* and *r/w*. One of each pair is normally present, giving rise to four possible determinant combinations and hence subtypes - *adr, adw, ayr* and *ayw* (Le Bouvier *et al.*, 1973). However, the cloned HBV DNA used in the work presented in this thesis is of a complex subtype *adyw*. The serum from which this virus was originally isolated displayed this serotype (Burrell *et al.*, 1979) and may have arisen *via* infection by more than one virus subtype.

While full immunogenicity of HBsAg is apparent in the context of a dimer of the molecule (Mishiro *et al.*, 1980), localisation of specific determinants, for example the *a* determinant (Bhatnagar *et al.*, 1982; Prince *et al.*, 1982) and, more generally, sites

within the S protein which elicit production of virus-neutralising antibodies (Gerin *et al.*, 1983; Purcell and Gerin, 1985) have been effected with the use of short peptides. Similar peptide studies involving the preS regions (Neurath *et al.*, 1985; 1986b; 1987; Iwarson *et al.*, 1985a; Milich and McLachlan, 1987; Emini *et al.*, 1989) taken together with studies involving the full-length proteins (Milich *et al.*, 1985; Schlicht *et al.*, 1987) demonstrate that the preS region, and in particular preS2, is more immunogenic than the S region. Indeed antibodies specific to the preS2 sequences are virus-neutralising (Iwarson *et al.*, 1985; Neurath *et al.*, 1986b; Emini *et al.*, 1989) and can also cross-react with HBV subtypes other than that against which they were raised (Neurath *et al.*, 1987; Emini *et al.*, 1989). These observations, when taken together with the finding that peptides corresponding to both the preS1 and preS2 regions can prime T cell help for antibody production to epitopes within preS1, preS2 and S, thereby potentially circumventing the non-responsiveness to vaccination with the S protein only (Milich *et al.*, 1985; Milich and McLachlan, 1987); has led to the suggestion that preS sequences be included in any future HBV vaccine.

1.4.2 Polymerase

The first indication of a HBV polymerase activity came in 1973 when it was discovered that preparations of HBV from the sera of chronically-infected individuals contained an endogenous polymerase activity capable of filling-in the incomplete strand of the HBV genome (Kaplan *et al.*, 1973). That the P gene is expressed during HBV infection has been demonstrated by the detection of polymerase-specific antibodies in sera from a convalescent chimpanzee (McGlynn and Murray, 1988). When the sequence of HBV became known (Galibert *et al.*, 1979; Pasek *et al.*, 1979) a candidate gene for production of this polymerase activity was identified, the P open reading frame, on the basis of the molecular weight of the predicted product. In addition, this sequence shares homology with the polymerase genes of retroviruses and other genetic elements which employ reverse transcription, in domains responsible for DNA polymerase/reverse transcriptase and RNaseH activities (Toh *et al.*, 1983; Khudyakov and Makhov, 1989; Radziwill *et al.*, 1990), thus proposing this gene

product as containing the activities required for replication of the HBV genome (See section 1.5). However, efforts to examine this protein have been hampered by its low abundance in HBV virions and its insoluble nature, and expression in tissue culture and heterologous systems had failed to produce protein which was sufficiently enzymatically active for detailed biochemical study. While sufficient, and sufficiently active, protein has been produced in such systems to indicate that the reverse transcriptase and RNaseH activities required for HBV replication reside within this protein (McGlynn and Murray, 1988; Bavand *et al.*, 1989; Chang *et al.*, 1990; Radziwill *et al.*, 1990), a recent report may allow more detailed biochemical analysis in the future. Wang and Seeger (1992) and Howe *et al.* (1992) report expression of the polymerase ORF of the duck hepatitis B virus in a cell-free rabbit reticulocyte lysate system, in an enzymatically active form, expressing reverse transcriptase activity, while McGlynn *et al.* (1992) report expression of enzymatically active HBV polymerase/reverse transcriptase in the baculovirus system. A more detailed discussion of the roles of the HBV P gene product is presented in Section 1.5.

Initial observations that a protein was tightly attached to the HBV genome (Kaplan *et al.*, 1973; Landers *et al.*, 1977) and specifically to the 5′ terminus of the long, minus strand (Gerlich and Robinson, 1980; Molnar-Kimber *et al.*, 1983) led to much speculation as to its nature and role. Evidence has now accumulated, both in the DHBV system (Bartenschlager and Schaller, 1988; Radziwill *et al.*, 1988; Bosch *et al.*, 1988; Wang and Seeger, 1992) and for HBV (Radziwill *et al.*, 1990) that the genome-linked protein is encoded by the P gene. It has further been demonstrated by the use of antibodies specific for different segments of the P polypeptide, that an N-terminal domain of the P gene product, distinct from its other functional domains, is responsible for this DNA linkage (Bosch *et al.*, 1988; Bartenschlager and Schaller, 1988) and indeed, amino acid sequence comparison of this area of the P protein with the terminal proteins of picornaviruses found a region of conservation with the VPg protein of human rhinovirus (Khudyakov and Makhov, 1989) leading the authors to suggest that the HBV protein acts in a similar manner to VPg; that is, as a protein primer for DNA synthesis. This has recently been confirmed experimentally by Wang

and Seeger (1992). However, this region alone is not sufficient to prime DNA synthesis, and neither, in the *in vitro* system used by the authors, is the full protein sufficient to polymerise more than a few nucleotides. This observation suggests that some modification of the protein may be required to alter its function from a primer to a polymerase, perhaps by formation of a dimer (Wang and Seeger, 1992) or even cleavage at the end of the terminal protein domain, thus releasing the reverse transcriptase, polymerase and RNaseH domains (Bartenschlager and Schaller, 1988).

1.4.3 X Antigen

Until recently, no biological function had been described for the product of the X ORF, hence its name. That it is expressed during HBV infection has been demonstrated by the presence of antibodies specific to HBxAg in the sera of HBV-infected individuals (Kay *et al.*, 1985; Moriarty *et al.*, 1985; Meyers *et al.*, 1986; Weber *et al.*, 1988). Detection of X-encoded protein, although with molecular weight slightly higher than expected, from liver tissue of HBV-infected individuals with X-specific antibodies has also been reported (Moriarty *et al.*, 1985), in addition to detection in human hepatoma cells with integrated HBV DNA (Moriarty *et al.*, 1985; Chisaka *et al.*, 1987) and in cell lines transiently transfected (Chang *et al.*, 1987) or transformed (Pugh *et al.*, 1986) with HBV DNA.

It had been suspected that the X gene encoded the genome-linked protein, but this role has now been assigned to the product of the polymerase gene (Section 1.4.2) and it has become apparent that HBxAg functions as an effective transcriptional transactivator. A huge range of target sequences of wide-ranging origins have been shown to be susceptible to this transactivation by HBxAg (for a recent comprehensive review see Rossner, 1992) including the HIV LTR, the SV40 early promoter/enhancer, cellular β-interferon promoter and the c-myc promoter. HBxAg has also been shown to transactivate HBV DNA sequences, specifically, in conjunction with enhancer I or II, the core promoter, the preS1 promoter, the preS2 promoter and the X promoter. The promiscuous nature of this transactivation suggests a more general method of

action than specific binding of HBxAg to a target DNA sequence and it has indeed been demonstrated that HBxAg does not bind to its target DNA (Seto *et al.*, 1990). In addition, only a subset of promoters capable of transactivation by HBxAg can be transactivated in any one cell line (Seto *et al.*, 1989; Rossner, 1992) suggesting that HBxAg acts *via* multiple cell type-specific transcription factors. This more general method of action would also concur with the ability of HBxAg to transactivate both RNA polymerase II and RNA polymerase III promoters (Aufiero and Schneider, 1990), and the absence of any obvious sequence motifs within the 154 amino acids of HBxAg (Colgrove *et al.*, 1989). Recently, however, three short regions of HBxAg (amino acids 46-52, 61-69 and 132-139) have been shown to be essential for its transactivation function while amino- and carboxy-terminal amino acids appear to be dispensable. Several candidate cellular transcription factors have been identified for interaction with HBxAg, the most notable being NF- KB (Twu *et al.*, 1989; Faktor and Shaul, 1990; Lucito and Schneider, 1992), AP-2 (Seto *et al.*, 1990) and TFIIIC (Aufiero and Schneider, 1990), the latter in the case of RNA polymerase III promoter transactivation. The details of the mechanism of this action were recently made more clear by a report indicating that HBxAg transactivation proceeds by a complex tumour promoter pathway (Kekule *et al.*, 1993). The presence of HBxAg causes elevation of the level of diacyl glycerol (by an as yet unknown mechanism), which then activates protein kinase C. The latter protein in turn activates a wide variety of transcription factors and this would go some way to explaining the wide variety of promoters transactivated by HBxAg.

It has been reported that HBxAg has intrinsic serine/threonine protein kinase activity (Wu *et al.*, 1990a). However, the protein shares no homology with, and is much smaller than, known protein kinases and demonstration of protein kinase activity specific to HBxAg has not been achieved in other laboratories (including Rossner, 1991). A recent report in which the X gene was expressed in HepG2 cells from a recombinant vaccinia virus found a fraction of the HBxAg produced to be phosphorylated, suggesting a possible method of regulation of its activity (Schek *et al.*, 1991). The authors also reported detection of a higher molecular weight HBxAg

species, comprising approximately 10% of the total HBxAg, which may be glycosylated, and found that all the HBxAg produced in general had a short half-life. If this accurately reflects the situation during natural infection, the unstable nature of HBxAg may contribute to the very low level of HBxAg apparently present.

An enigma of the X gene is its presence, and conservation of sequence (Summers *et al.*, 1978; Marion *et al.*, 1980) and function (Colgrove *et al.*, 1989) in mammalian hepadnaviruses, but apparent absence in avian hepadnaviruses (Mason *et al.*, 1980; Sprengel *et al.*, 1988). It has been suggested (Feitelson and Miller, 1988) that sequences related to the X gene are present in the core gene of duck and heron hepatitis B viruses, but to date no experimental evidence has emerged to demonstrate HBxAg-like functions for these viruses. In addition, while ground squirrels infected with mutant GSHV lacking a functional X gene failed to develop normal GSHV infection (Ganem and Varmus, 1987), transient transfection of X-minus HBV DNA into several liver cell lines resulted in apparently normal levels of viral proteins, replicative intermediates and virion export (Blum *et al.*, 1992). Therefore, it may be that the importance of the X gene product to the viral life cycle is different in GSHV and HBV, or that X is crucially involved in the initial events of infection, which were not investigated in the latter study. If the X protein product is not central to the viral life cycle, its role may be in the pathogenicity of hepadnavirus infection.

Finally, due to the properties of HBxAg outlined above, and the observation that X sequences are often found to be integrated into the host DNA in liver tumours (Miyaki *et al.*, 1986) it has been proposed that the X gene product may contribute to hepatocarcinogenesis (Zahn *et al.*, 1988; Koshy and Hofschneider, 1989; Kekule *et al.*, 1993).

1.4.4 Core antigen

The core antigen of HBV is encoded by the C ORF which spans nucleotides 1 to 549, and the antigen therefore comprises 183 amino acids. Its amino acid sequence is highly conserved, both between subtypes of HBV and between mammalian hepadnaviruses (Figure 6). The C gene product is a protein of approximately 21kD which polymerises to form the viral nucleocapsid, the inner structure of the virus which houses the HBV genome and polymerase (Figure 1.2). The nucleocapsid is a regular structure with icosahedral symmetry which consists of approximately 180 subunits, as demonstrated by examination under the electron microscope (Onodera *et al.*, 1982). The amino acid sequence of the core protein of DHBV (not shown), while broadly similar to HBV and the other mammalian hepadnaviruses, has several regions of amino acid insertions, which apparently are manifested as tiny spikes protruding from the surface of the nucleocapsid (Mason *et al.*, 1980). Despite this characteristic, and the larger size of the nucleocapsids, the structure is sufficiently similar for the study of DHBV nucleocapsids to be relevent, in general, to HBV.

HBcAg has proved most amenable to expression in heterologous systems, including tissue culture cell lines (Gough and Murray, 1982; Roossinck *et al.*, 1986; Roossinck and Siddiqui, 1987; Ohori and Matsuda, 1989), yeast (*Saccharomyces cerevisiae*) (Kniskern *et al.*, 1986; Miyanohara *et al.*, 1986; Imamura *et al.*, 1988), baculovirus (Hilditch *et al.*, 1990), *Acetobacter methanolicus* (Schroder *et al.*, 1991) and *Xenopus* oocytes (Zhou and Standring, 1991). However, possibly the simplest and most commonly-used expression system for HBcAg is *E. coli* (Burrell *et al.*, 1979; Edman *et al.*, 1981; Stahl *et al.*, 1982). This system can be utilised to express large quantities of HBcAg which assemble within the bacterium into nucleocapsid-like structures. These have physical properties apparently identical to nucleocapsids isolated from infected liver, and examination with the electron microscope shows these recombinant particles to be morphologically indistinguishable from those produced *in vivo* (Cohen and Richmond, 1982).

Figure 1.6

Amino acid sequence of HBcAg of HBV subtypes, WHV and GSHV, with identity sequence shown below. Note the complete conservation of amino acids 1-10, and of the four cysteine residues at amino acid positions 48, 61, 107 and 183 (bold).

HBV *adyw*	Pugh *et al.* (1986)
HBV *adr*	Ono *et al.* (1983)
HBV *adw*	Ono *et al.* (1983)
HBV *ayw*	Galibert *et al.* (1979)
HBV *ayr*	Okamoto *et al.* (1986)
HBV *adw2*	Valenzuela *et al.* (1980)
HBV *adr4*	Fujiyama *et al.* (1983)
WHV	Girones *et al.* (1989)
GSHV	Seeger *et al.* (1984).

```
            1                                                   50
HBVadyw     MDIDPYKEFG ATVELLSFLP SDFFPSVRDL LDTAAALYRD ALESPEHCSP
HBVadr      MDIDPYKEFG ASVELLSFLP SDFFPSIRDL LDTASALYRE ALESPEHCSP
HBVadw      MDIDPYKEFG ATVELLSFLP SDFFPSVRDL LDTASALYRE ALESPEHCSP
HBVayw      MDIDPYKEFG ATVELLSFLP SDFFPSVRDL LDTASALYRE ALESPEHCSP
HBVayr      MDIDPYKEFG ASVELLSFLP SDFFPSIRDL LDTASALYRE ALESPEHCSP
HBVadw2     MDIDPYKEFG ATVELLSFLP SDFFPSVRDL LDTASALYRE ALESPEHCSP
HBVadr4     MDIDPYKEFG ASVELLSFLP SDFFPSIRDL LDTASALYRE ALESPEHCSP
WHV         MDIDPYKEFG SSYQLLNFLP LDFFPDLNAL VDTATALYEE ELTGREHCSP
GSHV        MDIDPYKEFG SSYQLLNFLP LDFFPDLNAL VDTAAALYEE ELTGREHCSP

IDENTITY    MDIDPYKEFG ....LL.FLP .DFFP....L .DTA.ALY.E .L...EHCSP

            51                                                  100
HBVadyw     HHTALRQAIL CWGDLMTLAT WVGTNLEDPA SRDLVVSYVN TNVGLKFRQL
HBVadr      HHTALRQAIL CWGELMNLAT WVGSNLEDPA SRELVVSYVN VNMGLKIRQL
HBVadw      HHTALRQAIL CWGELMTLAT WVGNNLQDPA SRDLVVNYVN TNMGLKIRQL
HBVayw      HHTALRQAIL CWGELMTLAT WVGVNLEDPA SRDLVVSYVN TNMGLKFRQL
HBVayr      HHTALRQAIL CWGELMNLAT WVGSNLEDPA SRELVVSYVN VNMGLKIRQL
HBVadw2     HHTALRQAIL CWGELMTLAT WVGNNLEDPA SRDLVVNYVN TNVGLKIRQL
HBVadr4     HHTALRQAIL CWGELMNLAT WVGSNLEDPA SRELVVSYVN VNMGLKIRQL
WHV         HHTAIRQALV CWDELTKLIA WMSSNITSEQ VRTIIVNHVN DTWGLKVRQS
GSHV        HHTAIRQALV CWEELTRLIT WMSEN.TTEE VRRIIVDHVN NTWGLKVRQT

IDENTITY    HHTA.RQA.. CW..L..L.. W...N..... .R...V..VN ...GLK.RQ.

            101                                                 150
HBVadyw     LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
HBVadr      LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
HBVadw      LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
HBVayw      LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
HBVayr      LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
HBVadw2     LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
HBVadr4     LWFHISCLTF GRETVLEYLV SFGVWIRTPP AYRPPNAPIL STLPETTVVR
WHV         LWFHLSCLTF GQHTVQEFLV SFGVWIRTPA PYRPPNAPIL STLPEHTVIR
GSHV        LWFHLSCLTF GQHTVQEFLV SFGVWIRTPA PYRPPNAPIL STLPEHTVIR

IDENTITY    LWFH.SCLTF G..TV.E.LV SFGVWIRTP. .YRPPNAPIL STLPE.TV.R

            151                                         183
HBVadyw     RR.....GRS PRRRTPSPRR RRSQSPRRRR SQSRESQC
HBVadr      RR.....GRS PRRRTPSPRR RRSQSPRRRR SQSRESQC
HBVadw      R...RDRGRS PRRRTPSPRR RRSQSPRRRR SQSRESQC
HBVayw      R.....RGRS PRRRTPSPRR RRSQSPRRRR SQSRESQC
HBVayr      RR.....GRS PRRRTPSPRR RRSQSPRRRR SQSRESQC
HBVadw2     RR...DRGRS PRRRTPSPRR RRSPSPRRRR SQSRESQC
HBVadr4     R.....RGRS PRRRTPSPRR RRSQSPRRRR SQSRESQC
WHV         RRGGARASRS PRRRTPSPRR RRSQSPRRRR SQSPSANC
GSHV        RRGGSRAARS PRRRTPSPRR RRSQSPRRRR SQSPASNC

CONSENSUS   R.......RS PRRRTPSPRR RRS.SPRRRR SQS....C
```

The ease of expression of HBcAg in heterologous systems, and the ability to purify nucleocapsids from infected liver (Barker *et al.*, 1974; Hruska and Robinson, 1977) have allowed HBcAg to be studied in detail. The ease of expression in *E. coli* is partly due to the apparent lack of post-translational modification of HBcAg. However, HBcAg can exist as a phosphoprotein. Incubation of nucleocapsids, purified from infected liver, with ^{32}P resulted in phosphorylation of HBcAg on Serine residues (Albin and Robinson, 1980; Gerlich *et al.*, 1982; Feitelson *et al.*, 1982) and ^{32}P orthophosphate-labelling of cells transiently transfected with the C gene of HBV produced a similar result (Roossinck and Siddiqui, 1987). In the DHBV system, Schlicht *et al.* (1989a) showed the majority of this phosphorylation to be localised to the serine-rich carboxy-terminus of DHBcAg and this has subsequently been confirmed to be the case for HBcAg also (Machida *et al.*, 1991; Yeh and Ou, 1991). In addition, it has been noted that HBcAg produced in *E. coli* is not phosphorylated (Schlicht *et al.*, 1989a). As HBcAg produced in *E. coli* is not phosphorylated, but still forms nucleocapsid-like particles, it would appear that HBcAg phosphorylation is not a requirement for core particle formation. The function of this phosphorylation is currently unclear. Pugh *et al.* (1989) demonstrated that intracellular, non-enveloped DHBV core particles are heterogeneous with regard to phosphorylation status, and that phosphorylation is on the surface of the particles, while nucleocapsids isolated from mature DHBV virions are not phosphorylated. This observation led the authors to propose that de-phosphorylation of nucleocapsids may act as a signal of maturation of their viral genome and target them for packaging and export from the cell as mature virus.

It is well-documented that the protein kinase activity associated with the viral nucleocapsid is very tightly linked to the core particles and that they co-purify during standard purification (Albin and Robinson, 1980; Gerlich *et al.*, 1982). The identity of the protein responsible for this kinase activity is currently unknown. Due to the close association of the activity with the nucleocapsid, it is possible that HBcAg itself is responsible. However, the amino acid sequence of HBcAg has no homology with known serine/threonine kinases, and that DHBcAg produced in *E. coli* is neither

phosphorylated itself nor has kinase activity (Schlicht *et al.*, 1989a) suggests that the activity is not intrinsic to HBcAg. A controversial report (Wu *et al.*, 1990a) suggested that the source of kinase activity is HBxAg (Section 1.4.3) but other laboratories have failed to confirm this (including Rossner, 1991). As no ORF product of HBV bears any resemblance to known serine/threonine kinases, it may be that the nucleocapsid-associated kinase activity is derived from the host cell. This idea is supported by the observation that proteins consisting of the COOH-terminus 40 amino acids of HBcAg fused to human alpha-globin and human growth hormone were phosphorylated within HBcAg sequences as normal (Yeh and Ou, 1991). As this phosphorylation proceeded in the absence of other HBV gene products, and it is unlikely that the 40 amino acids of HBcAg in the fusion proteins contain the kinase activity, it is probable that the phosphorylation was mediated by cellular kinases, and that this may also be the case in normal viral infection.

Another notable feature of the COOH-terminus of HBcAg is its extreme arginine-richness (Figure 1.6) This protamine-like nature (Pasek *et al.*, 1979) suggests a nucleic-acid binding activity for the protein, which has indeed been demonstrated. Nucleocapsids isolated from HBV-infected liver have been shown to have DNA-binding activity (Petit and Pillot, 1985; Machida *et al.*, 1991) while more extensive studies with HBcAg produced in *E. coli* has demonstrated that the protein can bind both DNA and RNA (Matsuda *et al.*, 1988; Gallina *et al.*, 1989; Hatton *et al.*, 1992). It has also been proposed that this nucleic acid-binding is not completely non-specific. Nucleocapsids isolated from infected liver bound HBV DNA more efficiently than the unrelated pBR322 DNA (Petit and Pillot, 1985), while examination of RNA contained within core particles produced in *E. coli* showed an enrichment of core protein mRNA over other cellular mRNAs (Birnbaum and Nassal, 1990). However, it is not clear whether this is due merely to the large amount of HBcAg mRNA present in the cells. Mutational analysis of HBcAg has confirmed the proposed location of nucleic acid-binding activity to the COOH-terminus of the protein. While Gallina *et al.* (1989) detected no nucleic acid within particles formed in *E. coli* from core proteins lacking the carboxy-terminus region, HBeAg (essentially a non-particulate truncated HBcAg.

See section 1.4.5) isolated from the serum of an infected individual was shown to bind DNA to a similar extent to HBcAg produced in *E. coli* (Matsuda *et al.*, 1988); and transient transfection of C gene DNA lacking the arginine-rich region coding sequences resulted in production of core particles competent for pregenomic RNA packaging. However, an intermediate situation has also been found in which particles lacking the COOH-terminus, produced in either *E. coli* or by transient transfection of HuH7 cells, have been shown to contain a small but detectable amount of RNA (Birnbaum and Nassal, 1990; Hatton *et al.*, 1992; Nassal, 1992a) suggesting the presence of an additional nucleic acid-binding motif between amino acids 1 and 144 of HBcAg. On the basis of secondary structure predictions of HBcAg, this site has been proposed to lie between amino acids 100 and 120 of the protein (Matsuda *et al.*, 1988) but experimental evidence has not yet emerged to confirm this.

The discovery of a role for the DNA-binding activity of HBcAg was perhaps unexpected as the function of HBcAg had been thought to be limited to binding and hence encapsidation of the RNA pre-genome prior to reverse transcription. This DNA-binding activity may have a role in the negative regulation of the human β-interferon gene by HBcAg observed by Twu and Schloemer (1989), as may the two nuclear localisation signals demonstrated to be present within the arginine-rich domain of the protein (Yeh *et al.*, 1990; Eckhardt *et al.*, 1991). A recent detailed dissection of the nucleic acid-binding capabilities of the COOH-terminus of HBcAg produced in *E. coli* (Hatton *et al.*, 1992) has revealed four repeated sequence motifs. Repeat I (150-RRRDRGRS-157), the first 8 amino acids of the protamine-like region in the *adw* subtype studied, was shown to be sufficient for RNA packaging, as determined by the RNA content of purified core particles electrophoresed on agarose gels. In addition, this sequence could be replaced by four arginines with no significant alteration in RNA packaging capability resulting. While this sequence (Repeat I) did not bind DNA, the three other repeats bound DNA much more efficiently than RNA and the authors suggest that this function may be required during the process of reverse transcription, in order to make it energetically more favourable. This and other reports describing viruses with genomes mutant in the C gene to be deficient in various

aspects of genome replication (Schlicht *et al.*, 1989a; Nassal, 1992a) suggest a more extensive role for HBcAg in genome replication than was previously thought, and this will be discussed in more detail in Section 1.5.2.

That assembly of recombinant HBcAg into morphologically normal core particles proceeds spontaneously in the heterologous expression systems described above demonstrates that no other HBV gene products are required to catalyse the process, or to form part of the nucleocapsid structure itself. That similar particles are produced in systems as diverse as liver-derived cell lines, insect cells and *E. coli* also suggests that no specialised host proteins are involved. This has been confirmed by the observation that core particles produced in *E. coli*, and subsequently purified, can be dissociated into monomers which can then be re-formed into particles which are normal, by morphological and immunological criteria (M. Dyson, personal communication). In addition, the purification procedure used by Stahl and Murray (1989) involved the dissociation of particles formed from HBcAg fusion proteins, followed by re-association in the absence of contaminating proteins. Therefore, it would appear that the ability to form nucleocapsid structures is intrinsic to HBcAg and requires neither the viral components contained within it in the mature virus particle, nor any cellular proteins. Requirements for assembly-competence within HBcAg itself have recently been investigated and will also form part of the work presented in this thesis.

Several reports have indicated that the COOH-terminal region of HBcAg is not required for formation of normal core particles, when expressed both in *E. coli* (Stahl and Murray, 1989; Salfeld *et al.*, 1989; Gallina *et al.*, 1989; Birnbaum and Nassal, 1990) and in liver-derived cells (Nassal, 1992a). Similarly, deletion of 36 amino acids at the COOH-terminus of DHBcAg did not prevent formation of nucleocapsid-like particles in HepG2 cells (Schlicht *et al.*, 1989a). In addition, the entire arginine-rich region can be replaced by foreign sequences of varying length, without disruption of the ability to form core particles (Borisova *et al.*, 1989; Stahl and Murray, 1989; Schoedel *et al.*, 1992; A. Shiau, personal communication). As these particles are

highly immunogenic with respect to both HBcAg and the fused foreign epitopes (discussed below) and react with antibodies to the foreign epitope, those sequences must be present on the exterior of the particle. However, as the arginine-rich native COOH-terminus of HBcAg has been demonstrated to bind nucleic acid and has been proposed to have a role in replication of the genome, it would be expected that this region would be located on the interior of the capsid. This notion concurs with the observation that while the COOH-terminus of HBcAg produced in *Xenopus* oocytes is readily accessible to degradation by proteinase K when in a non-particulate form, it is apparently resistant to proteinase K when it is in the context of particles, suggesting an interior location (Zhou *et al.*, 1992). In addition, no major antibody binding sites of core particles have been localised to the arginine-rich region and fusion proteins in which foreign epitopes are inserted at the start of, and without deletion of, the arginine-rich tail form particles whose foreign epitopes are recognised by antibody (Borisova *et al.*, 1989). However, the whereabouts of the COOH-tail within this structure are unknown. Therefore it may be that the very unusual and basic nature of the native COOH-terminus causes it to be internal to the particle, while more normal peptide regions are exposed on the surface.

Sequences at the NH_2-terminus have also been found to be required for the formation of HBcAg which is immunologically normal. Stahl *et al.* (1982) demonstrated that HBcAg produced in *E. coli* which consisted of 8 amino acids of β-galactosidase followed by 3 linker amino acids followed by amino acids 3-183 of HBcAg reacted well with polyclonal anti-HBcAg while a protein which differed only in the deletion of amino acids 3 and 4 of HBcAg, failed to react significantly with antibody. Similarly, Salfeld *et al.* (1989) found a drastic loss of antigenicity with a protein consisting of amino acids 12-183 of HBcAg. However, in neither case was it determined whether core particles were formed. Further investigation of the results of Stahl *et al.* (1982) will be presented in chapter 3 of this thesis.

While HBcAg itself is detected only rarely in the serum of infected individuals outwith Dane particles (Chemello *et al.*, 1988), it does occur in the liver early in acute

infections, and antibodies to HBcAg are the first antibodies to appear following infection (Cohen, 1978) and are produced at a high titre (Figure 1.1). Trevisan *et al.* (1982) found that antibody bound to membranes of hepatocytes in liver from chronically infected patients is anti-HBcAg, and not anti-HBsAg, and Mondelli *et al.* (1982) demonstrated that T cells directed against hepatocytes of chronically-infected indivduals were directed specifically against HBcAg and not HBsAg. Therefore, it is apparent that HBcAg is displayed on the surface of infected hepatocytes, and this has been confirmed by immune electron microscopy of HBcAg in liver cell plasma membranes (Kojima *et al.*, 1987). Therefore an immune response against HBcAg may be important in the elimination of cells replicating virus. Also, as noted in section 1.1, the distribution of HBcAg is characteristic of the disease state of infected individuals, with focal distribution of HBcAg in livers of patients with chronic active hepatitis, and more diffuse distribution or absence in the more mild chronic persistent hepatitis state and cases of HBV infection with no apparent liver disease (Feitelson, 1989). This is concurrent with the suggestion that HBcAg may be recognised by the immune system, thereby contributing to HBV disease pathology by eliminating cells expressing HBcAg.

The major epitope on HBcAg recognised by anti-HBcAg antibodies has been located by Salfeld *et al.* (1989). Various segments of HBcAg were produced in *E. coli* and their cross-reaction with polyclonal and monoclonal antibodies examined, such that a single conformational determinant was mapped to a region around amino acid 80. As HBcAg is a non-secreted, internal component of the virion, it was perhaps surprising that antibodies directed against it are produced at such high levels during infection. However, this may be explained by the observation of Milich and McLachlan (1986) that HBcAg is both a T cell-independent and a T cell-dependent antigen. That is, it can stimulate B cells to produce anti-HBcAg antibodies both directly, as demonstrated by the induction of anti-HBcAg IgM and IgG in nude mice, despite the absence of T cells within these mice; and *via* T helper cells.

The T cell-independent stimulatory activity was observed only with particulate HBcAg

and non-particulate HBcAg required a helper T cell function to elicit antibodies. Investigations using inbred murine strains showed multiple but distinct T cell sites in HBcAg of 16-21 residues in size. The site recognised predominantly was dependent on the H-2 haplotype of the mouse examined (Milich and McLachlan, 1986), and further work showed that a synthetic peptide representing a T cell site was sufficient to prime production of antibody in a manner similar to the native protein (Milich *et al.*, 1987b). More surprising, however, was the observation in the same report that priming by a T cell site peptide or whole HBcAg followed by challenge with HBV elicited antibodies not only to HBcAg, but also to S, preS1 and preS2. This antibody production was achieved only if HBsAg and HBcAg were present in the same structure (the virion) and not if challenge was with a mixture of separate HBcAg and HBsAg particles. A model for the mechanism operating in this situation is dependent upon the observation that B cells can act as antigen-presenting cells (Lanzavecchia, 1985) and is depicted diagramatically in Figure 1.7. This model also explains the unexpected results of experiments in which chimpanzees were protected from infection by vaccination with HBcAg (Murray *et al.*, 1984; 1987; Iwarson *et al.*, 1985b; Tabor and Gerety, 1984). In addition, a mouse non-responsive to the S region of HBsAg produced anti-S antibodies after priming with a synthetic HBcAg peptide T cell site and challenge with virion, as anti-S antibody production could be mediated *via* a HBcAg-specific T-cell site and not the S T-cell site which was lacking in this mouse strain. Combined with the observation that almost 100% of chronically-infected patients produce high titres of anti-HBcAg (Hoofnagle *et al.*, 1973), the use of HBcAg as a vaccine against HBV is suggested as it appears to be effective in preventing HBV infection and could, perhaps, circumvent the non-responsiveness to the current HBsAg-containing vaccine, seen in 2-5% of vaccinees. The enhanced immunogenicity of HBcAg at both the B and T cell level also indicates a role as an effective carrier in multi-valent vaccines. To this end, very many fusion proteins have now been produced consisting of HBcAg sequences fused to a foreign epitope, either preS and/or S sequences of HBV (Stahl and Murray, 1989; Borisova *et al.*, 1989; Schoedel *et al.*, 1992) or epitopes from other viruses (Clarke *et al.*, 1987; Stahl and Murray, 1989; Borisova *et al.*, 1989; Francis and Clarke, 1989; Beesley *et al.*, 1990; Clarke *et al.*,

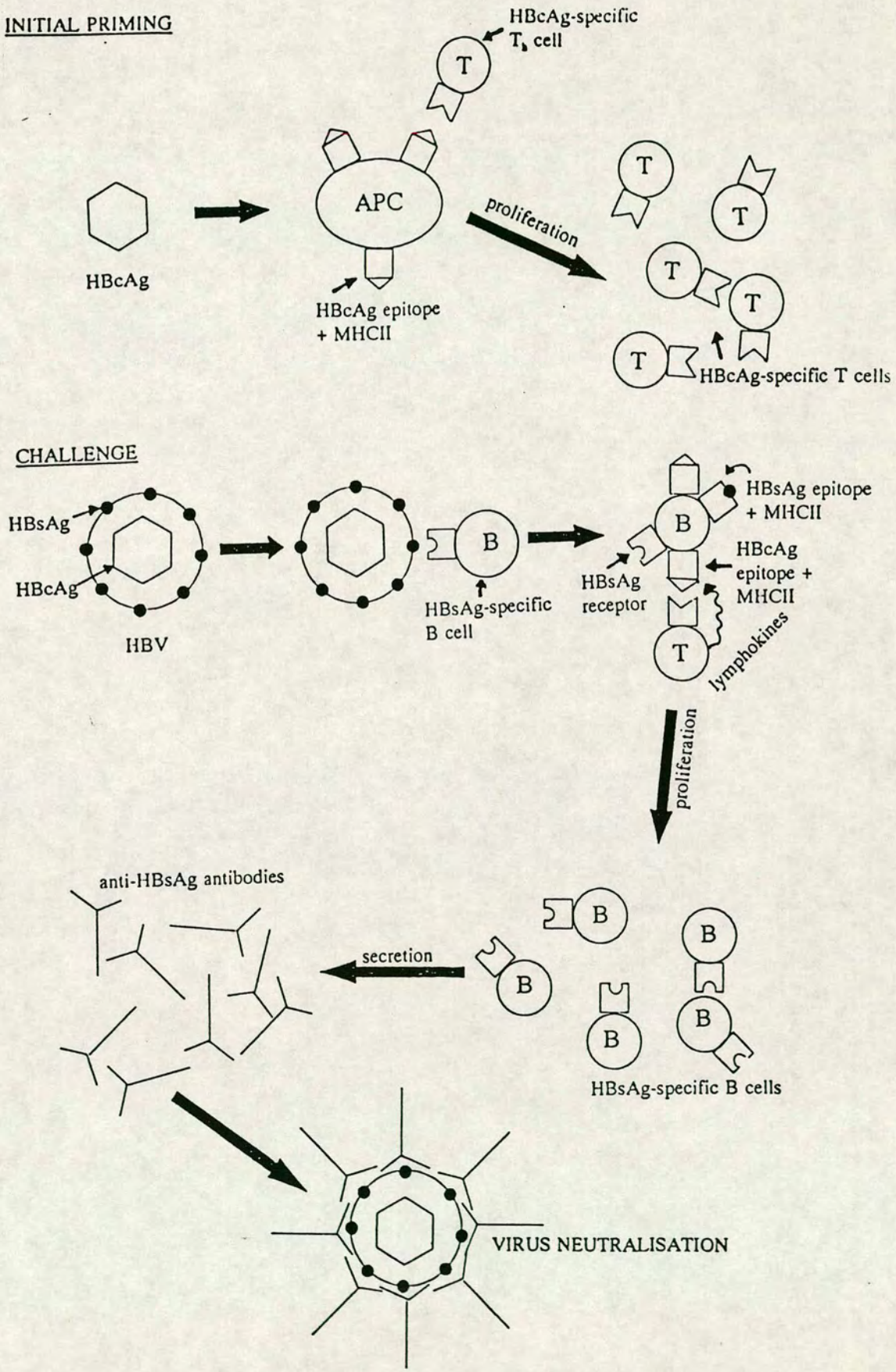
Figure 1.7

Initial Priming

HBcAg is taken up by an antigen-presenting cell (APC) which then displays HBcAg T cell epitopes, in conjunction with MHCII molecules, on its surface. These epitopes are recognised by a HBcAg-specific T cell which then proliferates to form an expanded population.

Challenge

Upon challenge with HBV, HBsAg eposed on the surface of the virus is recognised by a HBsAg-specific B cell. As this B cell can also act as an antigen-presenting cell (Lanzavecchia, 1985), it presents on its surface, in addition to its HBsAg receptor, both HBsAg and HBcAg peptides in conjunction with MHCII molecules. The HBcAg T cell epitopes within these peptides are then readily recognised by the expanded HBcAg-specific T cell population produced by the initial priming with HBcAg. Lymphokines released by such a T cell after recognition then stimulate the B cell, which is HBsAg-specific and therefore secretes HBsAg-specific antibodies, which bind to and neutralise the virus.


INITIAL PRIMING
HBcAg-specific
T$_h$ cell
T
APC
HBcAg
HBcAg epitope
+ MHCII
proliferation
T
T
T
T
T
HBcAg-specific T cells
CHALLENGE
HBsAg
HBcAg
HBV
B
HBsAg-specific
B cell
HBsAg epitope
+ MHCII
HBcAg
epitope +
MHCII
HBsAg
receptor
T
lymphokines
proliferation
anti-HBsAg antibodies
secretion
B
B
B
B
HBsAg-specific B cells
VIRUS NEUTRALISATION

1990; Francis *et al.*, 1990; Brown *et al.*, 1991; Urich *et al.*, 1991)

1.4.5 HBeAg

HBeAg was discovered in 1972 (Magnius and Espmark) in the serum of HBV-infected patients. At that time the authors suggested a link between the presence of HBeAg in serum and the "contagiousness" of the patient. This has since been shown to be the case and high titres of HBeAg are correlated with high levels of virus replication and high titres of Dane particles in the serum, and therefore high infectivity (Okada *et al.*, 1976; Ganem, 1982). Similarly, seroconversion to anti-HBeAg is indicative of clearance of the infection. Indeed, a difference of more than one million-fold was observed between sera containing HBeAg and sera containing anti-HBeAg when their infectivity was compared by titration in chimpanzees (Shikata *et al.*, 1977). Unlike HBcAg, HBeAg is found in the serum mainly as soluble dimers, a non-particulate structure, but it has also been demonstrated in association with Dane particles (Ohori *et al.*, 1979).

The source of HBeAg was determined when the conversion of HBcAg reactivity into HBeAg reactivity was achieved by treatment of purified nucleocapsids with pronase, sodium dodecyl sulphate, chaotropic agents or centrifugation in CsCl (Takahashi *et al.*, 1979; Budkowska *et al.*, 1979; Ohori *et al.*, 1980). However, an unequivocal demonstration that HBeAg could be generated from HBcAg was provided by MacKay *et al.* (1981) who converted HBcAg, produced in *E. coli* in the absence of other HBV gene products, to HBeAg by proteolytic degradation under dissociating conditions, thereby proving that no other HBV protein was the source of HBeAg. The nature of the conversion of the 21kD HBcAg to the 16kD HBeAg was further defined by Takahashi *et al.* (1983) who, *via* amino acid sequencing, determined the COOH-terminus of HBeAg to be 33-35 amino acids upstream of the COOH-terminus of HBcAg, suggesting therefore, that cleavage of these amino acids was involved in the conversion.

It has subsequently been discovered that this is the second of two processing steps, and that native HBcAg is not the natural precursor. Upstream of the initiation codon of the C gene lies another in-frame AUG which has been shown to be utilised *in vivo* (Section 1.3.1). Translation of this longer mRNA yields a core protein with an additional 29 amino acids at the NH_2-terminus. This additional amino acid sequence is termed the pre-core region which, due to its hydrophobic nature, was proposed to be a signal peptide (Sabatini, 1982) and this has proved to be the case. Transient transfection of COS cells with the C gene plus or minus pre-core sequences in an SV40 vector showed that the presence of the pre-core region resulted in secretion of HBeAg, and a strong association of p25, the HBeAg precursor protein, with cell membranes, probably the endoplasmic reticulum (Ou *et al.*, 1986). Similarly, partial deletion of pre-core sequences in stably-transformed rat cell lines resulted in the abolition of secretion of HBeAg (Roossnick *et al.*, 1986). The signal peptide also functions in *Xenopus* oocytes (Standring *et al.*, 1988) and while secretion signalled by a eukaryotic signal sequence would not be expected to occur in bacteria, expression of pre-C plus C in *E. coli* results in production of a protein with a strong affinity for the cellular membrane (Uy *et al.*, 1986). The observation that intracellular HBeAg is approximately 21kD, as opposed to the 24kD coding capacity of pre-C plus C is consistent with co-translational cleavage of a signal sequence by a signal peptidase (McLachlan *et al.*, 1987; Jean-Jean *et al.*, 1989a). In fact, it is only the first 19 amino acids of the 29 amino acid NH_2-terminal extension which form the signal sequence and are cleaved off, leaving 10 amino acids fused to the HBcAg polypeptide (Bruss and Gerlich, 1988; Garcia *et al.*, 1988; Standring *et al.*, 1988). A striking feature of the ER translocation of HBeAg is its inefficiency. Precursor proteins with the signal sequence still attached were detected consistently in the sera of HBV carriers (Takahashi *et al.*, 1991) and the livers of WHV-infected woodchucks (Weimer *et al.*, 1987), and *in vitro* assays have demonstrated abortion of translocation of pre-C, following signal sequence cleavage, in 70-80% of cases, followed by release of the product into the cytoplasm, in the *in vitro* translation/translocation system employed (Garcia *et al.*, 1988). This translation abortion may therefore be responsible for the protein populations described by McLachlan *et al.* (1987) and Jean-Jean *et al.* (1989)

(above). Similarly, in the *Xenopus* oocyte system a significant fraction of p25 precursor proteins evade translocation and processing (Yang *et al.*, 1992). Ou *et al.* (1989) report that such proteins can then be transported into the nucleus and this may be similar to the nuclear transport observed for HBcAg (Summers *et al.*, 1990). As the conformations of HBcAg and the p25 precursor proteins are different, it may be that removal of the signal sequence allows the correct configuration for nuclear transport to form. However, the role of the results of this inefficiency of HBeAg precursor translocation in the viral life cycle is unclear.

The remaining 10 amino acids attached to the NH_2-terminus of HBeAg have also recently been shown to be important. HBe proteins expressed from recombinant vaccinia virus vectors and secreted under control of the authentic signal sequence were compared to proteins fused to the un-related influenza hemagglutinin signal sequence. The hemagglutinin sequence produced secreted HBeAg which was a disulphide-linked dimer, capable of forming nucleocapsid-like particles with HBe and HBc antigenicity, while HBeAg with the authentic signal sequence was a monomer with only HBe antigenicity which did not form core particles (Schlicht and Wasenauer, 1991). Further examination showed that the 10 amino acids which remain linked to the HBcAg sequence are the most important factor in the difference in properties of HBeAg and HBcAg. In particular, the presence of the cysteine residue within this sequence prevents formation of disulphide-linked dimers (possibly by forming an intra-molecular disulphide bond) as a protein in which the cysteine is mutated forms such dimers, and hence particles, which have both HBe and HBc antigenicity (Wasenauer *et al*, 1992; Schodel *et al.*, 1993). Also, mutation of Trp-Leu-Trp, within the 10 amino acid peptide, to less hydrophobic amino acids allowed core particles to assemble, independent of the presence of the cysteine (Wasenauer *et al.*, 1992). Proteins consisting of the 29 amino acids of the pre-core region fused to HBcAg, when expresed in yeast, formed particles of 27nm diameter which were, however, unstable and readily dissociated by high salt (Miyanohara *et al.*, 1986). In this case, the additional 19 amino acids at the NH_2-terminus may be masking the preventive effect on particle formation of the 10 amino acids immediately upstream of the HBcAg Met.

After translocation to the ER, COOH-terminal cleavage of the protein takes place during travel through the secretory pathway (Standring *et al.*, 1988). Specifically, conversion of p22 to p16 is prevented by brefeldin A, which inhibits transport from the ER to the Golgi apparatus. Also, subcellular fractionation revealed p22 in both ER and the Golgi, but p16 was detected only in the Golgi fraction, thereby suggesting that the COOH-terminal processing of p22 to p16 takes place in the Golgi apparatus (Wang *et al.*, 1991)

The detection of regions of homology between known aspartyl proteases and the NH_2-terminal region of the HBcAg sequence (Miller, 1987) prompted suggestion that HBeAg may be generated from its precursor by proteolytic self-cleavage. However, mutation of the proposed enzymatic site did not prevent secretion of appropriately-processed HBeAg from transiently-transfected human and mouse cell lines (Jean-Jean *et al.*, 1989b; Nassal *et al.*, 1989). It is therefore more likely that the processing protease involved may be of cellular origin. Specifically, the observation that pepstatin both reduced HBeAg secretion and caused accumulation of the 22kD processing intermediate (Jean-Jean *et al.*, 1989b) suggests that the protease may be an aspartyl protease.

While HBeAg and HBcAg have a large degree of amino acid identity, they are serologically distinct. Monoclonal antibodies specific to HBeAg or HBcAg have been produced and with these it has been shown that while HBeAg, in its non-particulate form, has only HBe antigenicity conferred by two main epitopes; particulate HBcAg has both a HBcAg determinant and a HBeAg determinant. The HBe determinant shared by HBcAg and HBeAg is linear, while the determinants specific to HBcAg and HBeAg are conformational (Milich *et al.*, 1988; Salfeld *et al.*, 1989). Surprisingly, the linear HBe determinant overlaps with the conformational determinant but they are presented simultaneously, suggesting that the subunits of the nucleocapsid may be present in slightly differently-folded conformations. Thus it appears that while both HBeAg and HBcAg share a linear epitope, the HBc nucleocapsid structure exposes a conformational HBc-specific epitope determinant while masking the HBe-specific

determinant, and that the opposite situation is true for HBeAg (Salfeld *et al.*, 1989).

As the T cell epitopes identified by Milich *et al.* (1987a) consist of sequences common to HBcAg and HBeAg, it was expected that these antigens would be equivalently immunogenic at the level of T cell activation and this has been demonstrated to be the case (Milich *et al.*, 1988). In one strain of mice the T cell response to HBcAg was superior to that to HBeAg, suggesting the presence of a T cell epitope within the arginine-rich COOH-terminus of HBcAg, but such an epitope has yet to be identified (Milich *et al.*, 1988). This cross-reactivity at the T cell level may, in part, be responsible for the high levels of anti-HBcAg antibodies produced during infection, despite the low levels of HBcAg exposed to the immune system. The large amount of HBeAg in circulation in an infected individual may cause expansion of a T_h cell population which can drive anti-HBcAg production, in addition to anti-HBeAg. However, the levels of anti-HBe detected in serum are generally lower than those of anti-HBc and this may be due to both the lower efficiency of HBe in terms of stimulation of antibody production (Milich *et al.*, 1988), and the forming of immune complexes between HBeAg and anti-HBeAg, thus reducing the amount of "free" anti-HBeAg available for detection in standard assays.

The conservation of pre-core and core sequences between mammalian hepadnaviruses suggests an important function of HBeAg for the virus, but to date this function has not been determined. HBeAg produced from recombinant vaccinia viruses in HepG2 cells can be either secreted, or incorporated into the outer cell membrane where it is recognised by anti-HBe positive serum from HBV-infected patients (Schlicht and Schaller, 1989). If this situation is reflected *in vivo*, antibody-mediated elimination of such cells may occur, which would concur with the correspondence of seroconversion from HBeAg to anti-HBeAg with virus clearance. However, this would not be advantageous to the virus itself. In the DHBV system, mutation of the pre-C region had no effect on viral replication, assembly or infectivity (Chang *et al.*, 1987; Schlicht *et al.*, 1987; Schneider *et al.*, 1991) and so the reason for retention of the pre-C region in HBV, and HBeAg expression during viral infection, remains unknown.

1.5 HBV Replication

1.5.1 Receptor-Binding

The first stage in HBV infection is the binding of the virus to the hepatocyte. Until recently, the molecules involved in this reaction were very difficult to identify due to the lack of an *in vitro* infection system. However, the development of primary duck hepatocyte tissue culture (Tuttleman *et al.*, 1986b) and the achievement of infection of HepG2 cells with HBV virions obtained from the serum of a chronic carrier (Bchini *et al.*, 1990) may make this goal more easily attainable.

The observation that polymerised human serum albumin (pHSA) binds to HBV virions (Imai *et al.*, 1979) suggested that this molecule may act as an intermediate in HBV/receptor binding. It has subsequently been shown that pHSA binds to both the preS2 region of surface antigen (Machida *et al.*, 1983; 1984, Persing *et al.*, 1985) and the pHSA receptor present on hepatocytes. However, while the glutaraldehyde-polymerised pHSA used in the above assays binds the preS2 region well, this is not true for naturally-occurring pHSA, which does not bind significantly. In addition, human serum albumin occurs in sera naturally predominantly in a monomeric form and so the relevance of this observation to the *in vivo* situation is far from certain. However, several reports have confirmed that it is the preS region which is predominantly involved in hepatocyte binding. Assays involving affinity columns of cellulose-linked surface antigens through which were passed HepG2 cells or non-nuclear fractions of HepG2 cells, found that binding was inhibited by antibodies specific to the preS region and not the S region (Neurath *et al.*, 1985). Specifically, the preS1 region appears to have a critical role in hepatocyte-binding as antibodies to preS1 prevented binding, while the reduction, but not abolition, of binding achieved by anti-preS2 antibodies suggests an auxiliary role in binding for preS2 (Neurath *et al.*, 1986a). These results were mirrored by Pontisso *et al.* (1989) who employed both solid and liquid phase assays in which liver plasma membranes were exposed to the three forms of surface antigen. Only proteins containing the preS1 region could bind

directly to liver plasma membranes, but in the presence of pHSA, preS2/S proteins also bound, and binding of preS1/preS2/S proteins was enhanced. A recent report employing anti-idiotype antibodies has further implicated preS1 as the HBV polypeptide responsible for hepatocyte receptor-binding (Petit *et al.*, 1992).

While S sequences are highly conserved between hepadnaviruses this is not true for the preS regions. Therefore binding mediated by preS may explain the strict species-specificity of hepadnaviruses. There is also some sequence variability between human HBV subtypes but antibodies raised to synthetic peptides from preS1 and preS2 of one subtype of HBV were found to recognise all HBV subtypes (Neurath *et al.*, 1987). Therefore, antibodies raised to preS would prevent HBV binding to hepatocytes and thereby prevent infection. This would obviously be a useful response to a vaccine and it has been proposed that preS proteins should be included in future HBV vaccines. The results of Neurath *et al.* (1987) described above, demonstrate that the sequence variability of preS betwen different human isolates does not represent an impediment to the use of preS immunogens for worldwide immunisation.

Finally, the discovery by Peeples *et al.*, (1987) of a receptor for the S protein on Vero cells (derived from African Green Monkey kidney) has led to speculation that a similar receptor may exist on extrahepatic cells infected with HBV, but as yet there is no experimental evidence to confirm this.

Following uptake of the virion, the viral nucleocapsid is transported to the nucleus (Summers *et al.*, 1990) where viral genome replication begins.

1.5.2 Genome Replication

Examination of HBV DNA isolated from infected liver revealed, in addition to relaxed circular DNA, covalently-closed circular DNA and linear single-stranded DNA of various lengths, presumed to be nascent DNA (Ruiz-Opazo *et al.*, 1982; Mason *et al.*, 1982; Weiser *et al.*, 1983). The discovery that this single-stranded DNA was of the

(-) type only, and that no (+) single-stranded DNA was detected (Mason *et al.*, 1982) was inconsistent with a mechanism by which (-) DNA was produced by polymerisation on a (+) DNA template, and was the first indication that hepadnaviral replication proceeds by an unusual mechanism. A major step in the elucidation of this mechanism was made by Summers and Mason (1982) who demonstrated that the polymerase present within HBV nucleocapsids isolated from infected liver, synthesised the virus (-) strand DNA from an RNA template and degraded the RNA template as it was copied. Thus reverse transcription of an RNA intermediate is central to HBV genome replication. The full replication cycle has now been elucidated (reviewed by Seeger *et al.*, 1991) and is represented in Figure 1.8.

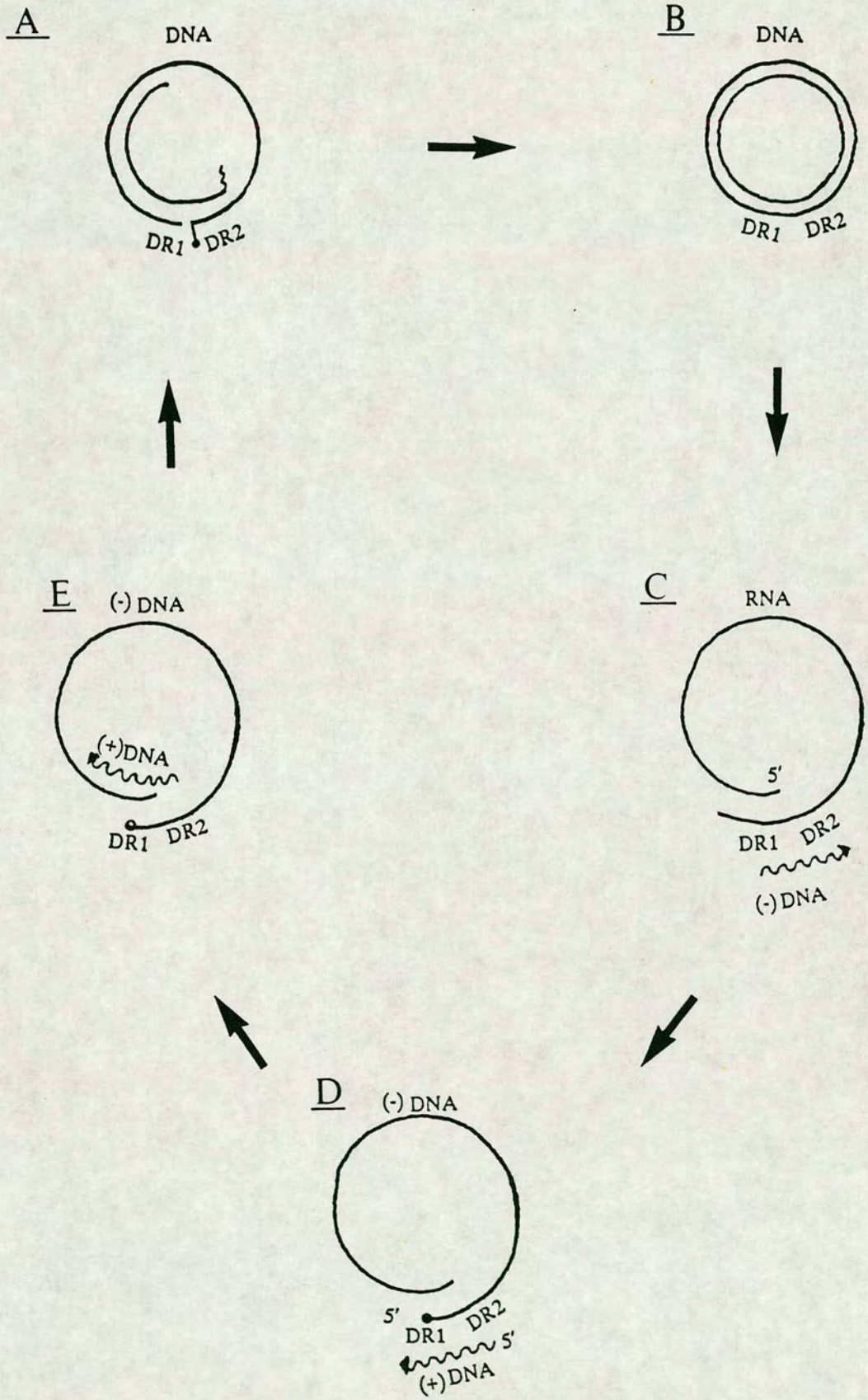
The first step in viral genome replication is the formation of covalently-closed circular DNA (cccDNA) from the partially double-stranded genome present in mature virions. cccDNA can be detected in the liver of DHBV-infected ducks 6-16 hours after infection (Mason *et al.*, 1983; Tagawa *et al.*, 1986), before other HBV DNA or RNA intermediates. Formation of cccDNA requires not only the filling in of the single-stranded gap but also the removal of the genome-linked protein, the removal of the terminal redundancy on the (-) strand DNA of the genome, followed by ligation of the 5′ and 3′ ends of this strand, the removal of the RNA primer present at the 5′ end of (+) DNA and the ligation of the 5′ and newly-formed 3′ end of (+) DNA. It is not yet clear whether these processes are facilitated by cellular or viral products, and while it has been demonstrated that the HBV polymerase can fill in the single-stranded gap in the genome (Kaplan *et al.*, 1973), specific inhibition of the viral enzyme by foscarnet does not prevent cccDNA formation (Mason *et al.*, 1987), suggesting that this function can also be carried out by a cellular polymerase activity. It is therefore unclear which enzyme completes (+)-strand DNA synthesis *in vivo*.

cccDNA is then amplified within the infected cell, in the absence of further infection (Tuttleman *et al.*, 1986a; Petcu *et al.*, 1988; Wu *et al.*, 1990b) before being utilised as the template for transcription of pre-genomic RNA (described in section 1.3.1). This terminally-redundant RNA contains three copies of an 11bp direct repeat (DR) which

Figure 1.8

Replication cycle of the HBV genome.

The partially double-stranded DNA genome (A) is converted to covalently-closed circular DNA (B) which is then amplified. Transcription begins upstream of DR1 and terminates on recognition, at the second encounter, of the polyadenylation signal downstream of DR1; thus forming the pre-genomic RNA (C). Reverse transcription of the RNA pre-genome initiates within DR1 and proceeds to the 5′ end of the template, thus forming (-) DNA with a short terminal redundancy and the genome-linked, polymerase-derived, protein attached to its 5′ terminus (D). The RNA template is degraded during this process. Plus-strand DNA synthesis initiates at DR2 and continues to the 5′ end of the (-) DNA strand (D). DNA synthesis continues after a template switch to the 3′ end of (-) DNA strand (E), to form the mature HBV genome.


A
DNA
DR1
DR2
B
DNA
DR1
DR2
C
RNA
5'
DR1
DR2
(-) DNA
D
(-) DNA
5'
DR1
DR2
5'
(+) DNA
E
(-) DNA
(+) DNA
DR1
DR2

contain signals important for minus- and plus-strand DNA synthesis (see below). Reverse transcription to synthesise the viral genome takes place in the nucleocapsid of the virus, and so the next stage in genome replication is the packaging of pre-genomic RNA into the nucleocapsid, proposed to occur by interaction of HBcAg with the pre-genomic RNA. Of all the RNAs produced by HBV, only genomic, and not sub-genomic, transcripts are packaged (Enders *et al.*, 1985). And of the genomic transcripts, only the shortest, which contain the initiation codon for core but not that for pre-core, are packaged (Will *et al.*, 1987; Enders *et al.*, 1987; Junker-Niepmann *et al.*, 1990). Both sets of transcripts contain the encapsidation signal within the pre-C sequence (Bartenschlager *et al.*, 1990; Junker-Niepmann *et al.*, 1990; Chiang *et al.*, 1992) but recognition of this signal in the context of the longer transcripts is believed to be blocked by the presence of ribosomes (Nassal *et al.*, 1990). The COOH-terminus of HBcAg, due to its protamine-like nature (Pasek *et al.*, 1979), was expected to be the region of the protein responsible for RNA encapsidation, and it has indeed been demonstrated that HBcAg lacking this region does not package nucleic acid in significant quantity (Gallina *et al.*, 1989). Deletion mutation of these sequences has localised this RNA encapsidation function to amino acids upstream of amino acid 164 (Nassal, 1992a) but the remainder of the COOH-terminus also appears to be required for other steps of genome replication, discussed below. In addition, the HBV polymerase gene product is required for RNA encapsidation (Bartenschlager *et al.*, 1990; Hirsch *et al.*, 1990; Chiang *et al.*, 1990; Chen *et al.*, 1992) and in fact binds to the encapsidation signal (Bartenschlager and Schaller, 1992), thereby explaining the very small amount of polymerase detected in virions.

The next step is reverse transcription of the RNA pre-genome by the viral polymerase gene product to form the (-) sense DNA strand of the genome, which has a terminal redundancy. Reverse transcription initiates in DR1 (Seeger and Maragos, 1990) and is primed by a protein primer, the reverse transcriptase itself, which remains covalently linked to the nascent DNA (Wang and Seeger, 1992). The RNA pre-genome is degraded during (-) DNA synthesis by the RNaseH activity of the viral polymerase (Radziwill *et al.*, 1990). The formation of the complementary (+) strand

DNA is also primed by an unusual mechanism. A capped oligoribonucleotide derived from the 5′ end of pregenomic RNA is found covalently attached to the 5′ end of (+) DNA (Dudley *et al.*, 1972; Lien *et al.*, 1986) and primes formation of (+) DNA, whose 5′ end maps to the first nucleotide downstream of DR2 (Lien *et al.*, 1986; Seeger *et al.*, 1986). However, as the (-) DNA template terminates just downstream of DR2, in order to form a complete (+) DNA strand, nascent (+) strands must switch to the 3′ portion of (-) DNA. This transfer may be facilitated by the genome-linked protein. Polymerisation then continues but not to completion, possibly due to lack of sufficient nucleotides within the nucleocapsid, hence the partial double-stranded nature of the genome. Core antigen has recently been implicated in (+) DNA synthesis by the observation that HBV with core antigen lacking amino acids 165-183 produces (+) DNA inefficiently. This may be due to a condensing function of HBcAg which would package DNA in a manner akin to histones, thus allowing (+) strand synthesis to proceed in the limited space inside the nucleocapsid (Nassal, 1992a).

1.5.3 Virion Assembly

As described in section 1.4.4, core particle assembly proceeds spontaneously, and once pre-genomic RNA and the attached polymerase is packaged and replication has begun, the particle can be considered to be a complete nucleocapsid. In order to become a complete virion this nucleocapsid must acquire an outer envelope. This is presumed to occur *via* an interaction between the nucleocapsid and HBsAg embedded in the endoplasmic reticulum, the primary position of HBsAg in the cell (Ou and Rutter, 1987). All three forms of surface protein are found in virions (Heerman *et al.*, 1984) and the L and S (and possibly M) proteins are essential for virion formation (Bruss and Ganem, 1991; Ueda *et al.*, 1991). Budding into the ER *via* this interaction results in the formation of Dane particles which are then secreted from the cell by normal pathways of vesicular transport and not by lysis of the infected cell.

It has been proposed that, as only replicated genomes (i.e. not pre-genomic RNA) are found in extracellular virions, genome replication causes an alteration in the

nucleocapsid, which then acts as a signal to trigger virus export. The discovery that HBcAg in mature DHBV virions is not phosphorylated, while intracellular HBcAg is phosphorylated (Pugh *et al.*, 1989), suggests that de-phosphorylation of HBcAg occurs during genome replication and may be such a signal.

1.6 Protein Structure and Disulphide Bonds

There are four levels of structure in proteins:

Primary structure	- the sequence of amino acids in the polypeptide chain
Secondary structure	- structural motifs such as alpha helices and beta sheets
Tertiary structure	- the formation of separate domains within a polypeptide chain, which contain elements of secondary structure.
Quaternary structure	- the association of separate polypeptide chains to form the functional protein.

HBV nucleocapsids contain all four levels of structure. As the rules for determining the structure of a protein are not as straightforward as those for DNA for example, protein structure prediction is very difficult and currently relies on computer programs which do not give wholly accurate results. Therefore experimental evidence for protein conformation is still required.

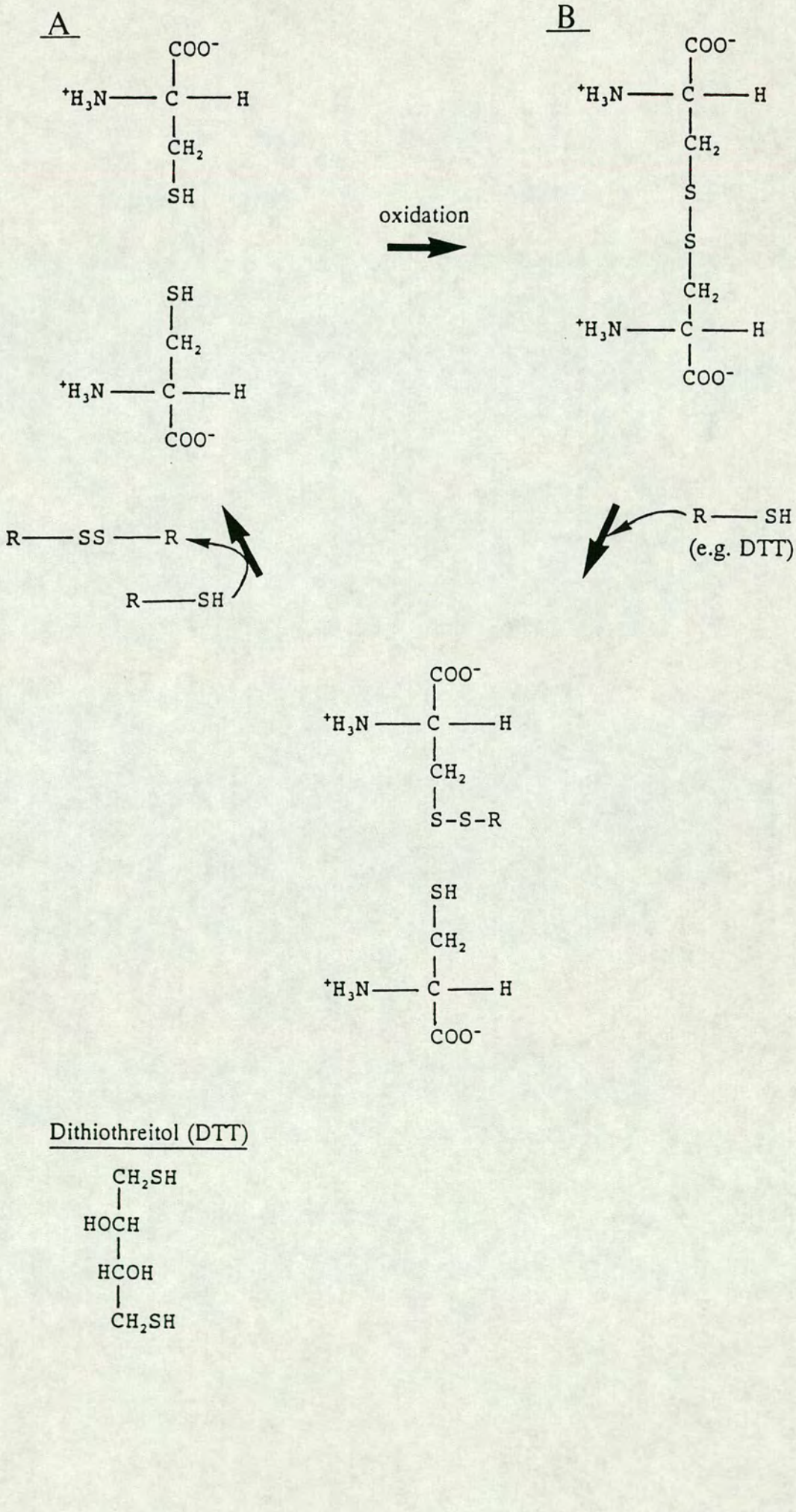
While the amino acids of a protein are covalently linked to form a chain, the areas of the chain brought together during folding are held together, in the main, by non-covalent forces (hydrogen bonds) which can be easily disrupted. For example, exposure to high temperature will cause the disruption of such hydrogen bonds, and the abolition of secondary, tertiary and quaternary structure. However, this will be impeded if disulphide bonds are present within the protein. These covalent bonds form between cysteine residues only, specifically between the sulphur atoms of these amino acids, as shown in Figure 1.9 (for a review of disulphide bond formation see Creighton, 1984a). The distance between the α-carbon atoms of the two cysteines

Figure 1.9

Disulphide bonds.

A disulphide bond (B) will form between two cysteine residues (A) which are in the correct positions, in an oxidative environment.
This bond can be broken by the action of an excess of reducing agent (R-SH) to create once again the two constituent cysteine residues (A).

The structure of dithiothreitol, the reducing agent used in this work, is shown at the bottom of the figure.



Dithiothreitol (DTT)

involved should be in the range 0.4-0.9ηm (Creighton, 1984b). These bonds are stable once formed, unless exposed to reducing agent which would result in the reversible cleavage of the bond (Figure 1.9), or by oxidants which can cleave disulphide bonds irreversibly, to the sulphinic (RSO_2H) and sulphonic (RSO_3H) acids. Other nucleophiles can also break disulphide bonds, reversibly. Therefore the nature of the environment in which cysteine residues are present is also important in determining whether disulphide bond formation will occur. As intracellular proteins are in a reductive environment, they generally do not contain disulphide bonds, but such bonds are often found in extracellular proteins. Extracellular proteins which have been secreted from cells have passed through the lumen of the endoplasmic reticulum, the first compartment of the secretory pathway, and it is thought to be here that disulphide bond formation occurs, possibly by the action of the enzyme protein disulphide isomerase which is localised to the ER (reviewed by Freedman, 1984). However, HBV core particles produced in *E. coli* form disulphide bonds only after disruption of the host bacteria and exposure to an oxidative environment, as they are not naturally secreted from the cell.

HBcAg has four cysteine residues which are completely conserved among mammalian hepadnaviruses and therefore all four are candidates for disulphide bond participation. As HBcAg molecules associate to form the nucleocapsid, the possibility exists for both intra- and inter-molecular disulphide bonds to be formed, and an investigation of disulphide bond formation in HBV core particles is presented in Chapter 4.

1.7 Aims of the Thesis

At the time when the work presented in this thesis was initiated, very little was known about the structure of HBV nucleocapsids. Electron microscopy had demonstrated that these nucleocapsids are regular structures of approximately 27 nm diameter, with icosahedral symmetry, and are composed of 180 subunits. As HBcAg has proved to be resistant to crystallisation, no crystal structure data was available and so molecular biological methods were required to investigate the structure further.

It had been demonstrated previously that when the C gene of HBV was expressed in *E. coli*, nucleocapsid-like particles were formed with morphology indistinguishable from nucleocapsids isolated from HBV-infected individuals, and so *E. coli* was an ideal system in which to produce core particles for further study. While creating plasmids from which higher levels of HBcAg could be expressed in *E. coli*, Stahl *et al.* (1982) found that expression of a plasmid encoding amino acids 3 to 183 of HBcAg resulted in an ammonium sulphate-precipitated extract which reacted well with anti-HBcAg antibodies, but that this was not true for a plasmid encoding amino acids 5 to 183 of HBcAg, which failed to react significantly with antibody. The aim of the work presented in Chapter 3 of this thesis was to investigate the reason for this lack of reactivity. This was to be achieved by creating a series of substitution and deletion mutants involving amino acids 3 to 6 of HBcAg and investigating their levels of expression in *E. coli*, and the reaction of the purified proteins with anti-HBc antibodies, alongside the original proteins.

The cloning and sequencing of various HBV subtypes and mammalian hepadnaviruses revealed complete conservation of the four cysteine residues of HBcAg. In combination with the unpublished observations of our laboratory that core particles would enter a denaturing polyacrylamide gel during electrophoresis only after incubation with an excess of reducing agent, this suggested that the cysteines of HBcAg may be involved in inter-molecular disulphide bonds. The aim of the work presented in Chapter 4 of this thesis was to determine, firstly, which if any of the cysteines of HBcAg are linked by inter-molecular disulphide bonds; secondly, which if any of the cysteines are involved in intra-molecular disulphide bonds, and thirdly to determine the importance of the completely conserved cysteine residues to the integrity and properties of the core particles. These aims were to be achieved by creation and examination of the behaviour during non-reducing SDS-PAGE of a panel of mutant proteins in which cysteines were substituted by serines in various combinations. Work by Stahl and Murray (1989) demonstrated that the removal of the arginine-rich carboxy-terminal region of HBcAg did not affect the ability of the protein to form nucleocapsid-like structures when expressed in *E. coli*. In order to

investigate whether removal of the carboxy-terminal region of the protein had any effect on disulphide bond formation in the particle, the cysteine mutants were also examined in the context of truncated proteins.

Finally, the aim of the work presented in Chapter 5 was to extend the investigation of the effect of the carboxy-terminus of HBcAg on disulphide bond formation in the remainder of the protein by examining disulphide bond formation in three HBcAg fusion proteinsby the use of non-reducing SDS-PAGE. These proteins, created by Stahl and Murray (1989), consist of foreign sequences fused to amino acid 144 of HBcAg, and the foreign sequences differ from each other, and from the carboxy-terminal region of HBcAg, in terms of both length and cysteine content.

While the specific aim of this work was to determine the role of particular amino acids of HBcAg in determining the structure of core particles, it was also anticipated that such observations could be a means of drawing further, more general, conclusions about the structure of the HBcAg monomer and of the particles it forms, in particular by indicating their spatial arrangements.

CHAPTER 2: Materials and Methods

2A MATERIALS

2A.1 SUPPLIERS OF LABORATORY REAGENTS

Restriction endonucleases:

Boehringer Mannheim plc; Mannheim, Germany

New England Biolabs Inc; Beverly, Massachusetts, USA

Pharmacia LKB Biotechnology; Milton Keynes, UK

E. coli DNA Polymerase I (Klenow fragment):

Boehringer Mannheim plc

Northumbria Biologicals Limited, Northumberland, UK

T4 DNA Ligase:

Boehringer Mannheim plc

T4 Polynucleotide Kinase:

Prepared by S. Bruce (Reference: Midgely and Murray, 1985)

Deoxyribonuclease I:

Sigma Chemical Co.; Poole, UK

Ribonuclease A:

Sigma Chemical Co.

M-MLV Reverse Transcriptase:

Gibco-BRL Life Technologies; Paisley, UK

Thermus aquaticus (Taq) Polymerase:

International Biotechnologies Inc. (IBI); New Haven, Connecticut, USA

Boehringer Mannheim plc

Deoxynucleoside triphosphates and dideoxynucleoside triphosphates:

Boehringer Mannheim plc

Radioactive nucleoside triphosphates and deoxynucleoside triphosphates:

Amersham International plc; Aylesbury, UK

Standard Laboratory Reagents (various grades):

BDH Chemicals Ltd; Poole, UK

Fisons Chemicals; Loughborough, UK

Gibco BRL Life Technologies

Sigma Chemical Co.

Bacterial Media Reagents:

Becton-Dickinson UK Ltd; Oxford, UK

Difco Laboratories; East Mosely, UK

2A.2 MICROBIOLOGICAL MATERIALS

2A.2.1 BACTERIAL STRAINS

Escherichia coli K12 strains used in this work are listed below:

Strain:	TG1
Relevant Genotype:	*hsd5, (lac,pro), supE, thi, [F' traD36, proA*$^{+B}$*+, lacZ M15, lacI*^{q}]
Use:	host for bacteriophage M13 vectors
Reference:	Amersham International (in catalogue supplied with product RPN2322)

Strain:	BMH71-18 *mut*L
Relevant Genotype:	*mutL*::Tn10, *(lac, pro), supE, thi, [F' proA*$^{+}$B^{+}*, lacZ M15, lacI*^{q}]
Use:	transformation host for heteroduplexes produced by SDM using bacteriophage M13
Reference:	Kramer *et al* (1984)

Strain:	W3110Iq
Relevant Genotype:	F' *lac* with *lacIq*
Use:	host for *tac* expression vectors
Reference:	Stahl and Murray (1989)

2A.2.2 BACTERIOPHAGE AND PLASMIDS

E. Coli plasmids and bacteriophage used in this work are listed below. Construction of derivative plasmids and bacteriophage is described in chapters 3 and 4.

Name: M13mp18

Description: *E. coli* bacteriophage M13-based vector used for generating single-stranded DNA for sequence determination and site-directed mutagenesis. This vector contains a portion of the *lacZ* gene and multiple cloning sites.

Reference: Messing and Vieira (1982); Norrander *et al.* (1983)

Name: pRI-II

Description: Amp^r. This plasmid contains the coding sequence for amino acids 1-8 of β-galactosidase followed by three linker amino acids (GluPheHis) followed by amino acids 3-183 of HBcAg, under the control of the *lac UV5* promoter, in a pBR322 background.

Reference: Stahl *et al.* (1982)

Name: pR1-4

Description: Amp^r. This plasmid contains the coding sequence for amino acids 1-8 of β-galactosidase folowed by three linker amino acids (GluPheHis) followed by amino acids 5-183 of HBcAg, under the control of the *lacUV5* promoter, in a pBR322 background.

Reference: Stahl *et al.* (1982)

Name: pTacHpaIIR2

Description: Amp^r, Tet^r. This plasmid contains the coding sequence for amino acids 1-8 of β-galactosidase followed by three linker amino acids (GluPheHis) followed by amino acids 3-144 of HBcAg, under the control of the *tac* promoter, in a pBR322 background

Reference: Stahl and Murray (1989)

Name: pHBcpres1(1-20), pHBcpres1(1-36), pHBcE46

Description: Amp^r. These plasmids are derived from pTacHpaIIR2 and have sequences inserted at the appropriate position to encode the following fusion proteins: β-galactosidase (amino acids 1-8) /GluPheHis/ HBcAg (amino acids 3-144) followed by, respectively:

HBV preS1 (amino acids 1-20)

HBV preS1 (amino acids 1-36)

HIV env (amino acids 728-751)

Reference: Stahl and Murray (1989)

2A.2.3 MICROBIOLOGICAL MEDIA

Luria Broth (L-Broth)

1% (w/v) Difco Bacto Tryptone

0.5% (w/v) Difco Bacto Yeast Extract

1% (w/v) NaCl

pH adjusted to 7.2

L-Agar

1% (w/v) Difco Bacto Tryptone

0.5% (w/v) Difco Bacto Yeast Extract

1% (w/v) NaCl

1.5% (w/v) Difco Agar

pH adjusted to 7.2

Top Agar

1% (w/v) Baltimore Biological Laboratories trypticase

0.5% (w/v) NaCl

1% (w/v) Difco agar

(0.1mM X-gal and 1mM IPTG if required)

5x Spizizen Salts

0.2% (w/v) $(NH_4)_2SO_4$

1.4% (w/v) K_2HPO_4

0.6% (w/v) KH_2PO_4

0.1% (w/v) tri-sodium citrate

0.02% (w/v) $MgSO_4$

Minimal Agar

1.5% (w/v) Difco Bacto Agar

0.2% (w/v) glucose

1x Spizizen Salts

Antibiotics

ampicillin used at 100μg/ml

2B GENERAL METHODS

2B.1 NUCLEIC ACID METHODS

2B.1.1 Phenol Extraction of Nucleic Acid

Phenol was equilibrated with 1M Tris-HCl pH8.0 prior to use. An equal volume of phenol was added to the solution of nucleic acid, vortexed thoroughly and subjected to centrifugation at 17000x g for 5 minutes at room temperature. The upper aqueous phase containing nucleic acid was removed and retained.

2B.1.2 Phenol/Chloroform/Isoamyl Alcohol (PCI), Chloroform or Butanol Extraction of Nucleic Acid

PCI	50% phenol
	48% chloroform
	2% isoamyl alcohol

These methods are the same as that for phenol extraction of nucleic acid but with PCI, chloroform or butanol substituted for phenol.

2B.1.3 Ethanol Precipitation of Nucleic Acid

TE	25mM Tris
	10mM EDTA
	pH adjusted to 8.0

To a solution of nucleic acid was added 0.1 volume of 3M sodium acetate pH5.2 and 2.5 volumes of cold (-20°C) absolute ethanol. This was mixed thoroughly and incubated at -20°C for a minimum of 30 minutes. Nucleic acid was sedimented by

centrifugation at 17000x g in a microfuge for 15-30 minutes, or at 12000x g in a Sorvall centrifuge for 30-45 minutes at 4°C. The pellet was dried under vacuum and redissolved in TE.

2B.1.4 Isopropanol Precipitation of Nucleic Acid

The method is that for ethanol precipitation of nucleic acid but substituting 0.4 volume 5M ammonium acetate for sodium acetate, and 2 volumes of cold (-20°C) isopropanol for ethanol.

2B.1.5 Quantification of DNA

The optical density (OD) of nucleic acid solutions at 260 ηm was measured using a Unicam Ultraviolet and Visible Spectrophotometer (SP500 Series 2). An OD_{260} value of 1.0 represents a concentration of 50μg/ml for double-stranded DNA and 38μg/ml for single-stranded DNA.

2B.2 ENZYMATIC MANIPULATION OF DNA

2B.2.1 DNA ligation

10x Ligation Buffer	500mM Tris pH 7.8
	100mM $MgCl_2$
	200mM DTT
	10mM ATP
	500μg/ml BSA

Ligation of cohesive ends was carried out with an insert:vector ratio of approximately 3:1. Ligation of blunt ends was carried out with an insert:vector ration of approximately 5:1. The DNA was incubated for 15-19 hours at 14°C in 1x Ligation

Buffer in the presence of 1 unit of T4 DNA ligase per μg of DNA.

2B.2.2 Restriction enzyme digestion

DNA was digested at 37°C, unless stated otherwise, in 1x restriction buffer (supplied by the manufacturer) by 5-10 units of enzyme per μg DNA. Reactions were terminated by either incubation at 70°C for 15 minutes or, when *Bam*HI was used, by phenol extraction followed by ethanol precipitaion.

2B.3 PURIFICATION OF DNA

2B.3.1 Electrophoresis of Nucleic Acid

10 x TBE	108g Tris 9.3g EDTA 55g Boric acid made to 1 litre with H_2O pH 8.3
5 x DNA Sample Buffer	15% (w/v) ficoll, molecular weight 400000 50mM EDTA 0.125% (w/v) bromophenol blue 0.125% (w/v) xylene cyanol FF

DNA was fractionated on the basis of fragment size by electrophoresis in submerged, horizontal, agarose slab gels. Agarose gels consisted of 1% (w/v) agarose, which was melted in 1 x TBE buffer in a microwave oven. Gels were cast containing 0.5μg/ml ethidium bromide. Prior to loading, 0.2 volume of 5 x DNA Sample Buffer was added to each sample. Electrophoresis was carried out in 1 x TBE buffer, at 5-20 volts/cm.

1kb Ladder (Bethesda Research Laboratories) was used for DNA size markers and consists of DNA fragments of sizes (bp):

12,216	8,144	4,072	1,018	298	142
11,198	7,126	3,054	516/506	220	75
10,180	6,108	2,036	394	200	
9,162	5,090	1,635	344	154	

2B.3.2 Warm Phenol method for elution of DNA from a low-melting point agarose gel

DNA was electrophoresed through a 1% low melting point agarose gel in the presence of 1μg/ml ethidium bromide. After viewing under an ultra-violet lamp, the piece of gel containing the appropriate fragment of DNA was excised with a clean scalpel. To this was added an equal volume of TE followed by incubation at 65°C until the agarose melted. This was then extracted twice with phenol (equilibrated to 37°C), once with PCI (room temperature) and once with butanol (room temperature). The DNA was recovered by ethanol precipitaion, dried under vacuum and resuspended in an appropriate volume of TE.

2B.3.3 "Geneclean II" (Bio 101 Inc) method for elution of DNA from an agarose gel

DNA was electrophoresed through a low melting-point agarose gel in 1x TBE. The appropriate piece of the gel was removed and to it was added 0.5 volume "TBE Modifier" (a mixture of concentrated salts) and 4.5 volumes 6M NaI followed by incubation at 50°C for 5 minutes. "Glassmilk" (a suspension of silica matrix in H_2O) was then added (5μl for 5μg or less of DNA and an additional 1μl for every further 0.5μg of DNA) and incubated on ice for 5 minutes (or 15 minutes if the total volume was greater than 1.5ml), with occasional mixing. The "Glassmilk"/DNA complex was

then pelleted by centrifugation at 17000x g at room temperature for 5 seconds and all supernatant removed. The pellet was washed 3 times with "New Wash" (50% ethanol, NaCl, Tris-HCl, Tris-base, EDTA pH 7.0-8.5), followed by addition to the pellet of an equal volume of TE. This was incubated at 50°C for 3 minutes and then subjected to centrifugation at 17000x g for 30 seconds at room temperature. The supernatant was removed and retained.

2B.3.4 Small-Scale Preparation of Plasmid DNA from *E. coli*

Lysis Buffer	25mM Tris-HCl pH8.0
	10mM EDTA
	50mM glucose
Lysis Solution	0.2M NaOH
	1% (w/v) SDS
	(freshly prepared)

1.5ml of a stationary phase *E. coli* culture, grown in the presence of ampicillin, was centrifuged at 17000x g for 2 minutes at room temperature. The supernatant was discarded and the cell pellet resuspended in 100μl of Lysis Buffer. 200μl of Lysis Solution was then added, mixed by inversion and incubated on ice for 5 minutes. After addition of 150μl of 3M sodium acetate pH 5.2 and incubation on ice for 20 minutes, the heavy precipitate of denatured protein and chromosomal DNA was pelleted by centrifugation at 17000x g for 10 minutes at room temperature. This centrifugation step was repeated to remove residual debris. The plasmid DNA was recovered from the supernatant by ethanol precipitaion and the resulting DNA pellet was washed with 70% (v/v) ethanol, dried under vacuum and resuspended in 50μl of H_2O.

2B.3.5 Small-Scale Preparation of M13 Replicative Form DNA from *E. coli*

1ml of L-Broth was innoculated with 15μl of a stationary phase culture of TG1 cells, and with a single fresh M13 plaque. After shaking for 6 hours at 37°C, the procedure followed was that for the small-scale preparation of plasmid DNA from *E. coli* (above).

2B.3.6 Large-Scale Preparation of Plasmids from *E. coli*

50ml of a stationary phase *E. coli* culture, grown in the presence of ampicillin, was centrifuged at 10000x g for 10 minutes at 4°C and the cell pellet resuspended in 3.5ml of Lysis Buffer. To this was added 8mg of lysozyme dissolved in 0.5ml Lysis Buffer and the mixture incubated on ice for 1 hour. 8 ml of Lysis Solution was added and thoroughly mixed, followed by incubation on ice for 1 hour. Addition of 5ml of 3M sodium acetate pH5.2 and vigorous mixing followed by incubation on ice for 1 hour produced a heavy precipitate which was pelleted by centrifugation at 11000x g for 45 minutes at 4°C. The supernatant was extracted with PCI and the DNA in the aqueous layer was recovered by precipitation with isopropanol and centrifugation at 11000x g for 15 minutes at 4°C. The DNA pellet was dried under vacuum and resuspended in 0.5ml of H_2O. Ribonuclease A was added, to a final concentration of 0.2mg/ml, followed by incubation at 37°C for 20 minutes. This was extracted with PCI until the interface between the aqueous and organic phases was clear and the DNA was then recovered from the aqueous phase by ethanol precipitation. The pellet was resuspended in TE to a final concentration of 1μg/μl.

2B.4 RNA METHODS

2B.4.1 DEPC-H_2O Preparation

To 100ml of H_2O was added 200μl of DEPC. This was shaken vigorously and allowed to stand at room temperature for 20 minutes before being sterilised in an autoclave.

All solutions used in RNA work were made using DEPC-H_2O.

2B.4.2 Small-Scale Preparation of RNA from *E. coli*

The method used was essentially that of Barry *et al.* (1992).

50μl of a stationary phase culture of the appropriate plasmid-carrying cells, grown in selecting medium, were added to 5ml L-Broth containing ampicillin and shaken at 37°C until OD_{650} = 1.0. IPTG was then added to a final concentration of 0.5mM and the culture shaken at 37°C for approximately 14 hours. 1.5ml of this culture was then centrifuged at 17000x g for 2 minutes, the supernatant discarded and the cell pellet resuspended in 20μl of DEPC-H_2O to which was added 3μl of 0.5% DEPC. 200μl ice-cold acetone was then added and mixed by hand before centrifugation at 17000x g for 2 minutes. The supernatant was removed and the pellet resuspended in 30μl of DEPC-H_2O. 1μl of 100μg/ml Proteinase K was added, the mixture incubated on ice for 10 minutes and 3.5μl DEPC (0.5%), 200μl phenol (pre-heated to 70°C) and 150μl chloroform added and mixed by hand. After centrifugation at 17000x g for 5 minutes the aqueous phase was removed and added to 1ml of ethanol and the RNA allowed to precipitate at -20°C for 1 hour before recovery by centrifugation at 12000x g for 15 minutes and resuspension of the RNA pellet in 100μl DEPC-H_2O. RNA was stored at -70°C.

2B.4.3 Removal of DNA from RNA Preparations

50μl of RNA from the above preparation was precipitated with ethanol and resuspended in 50μl of 10mM Tris/ 1mM EDTA. To this was added 50μl of DNase Cocktail and the mixture incubated at 37°C for 15 minutes. This reaction was stopped by the addition of 25μl of DNase Stop Mix followed by one PCI extraction, one chloroform extraction and ethanol precipitation and the resulting pellet resuspended in 50μl of DEPC-H_2O.

2B.4.4 RNase Treatment of Control RNA Samples

To 25μl of DNase-treated RNA was added 2μl of 10mg/ml RNase A which was then incubated at 37°C for 30 minutes. This was then extracted once with PCI and once with chloroform and then added to 2.5 volumes ethanol. After incubation at -20°C for 1 hour the solution was centrifuged at 12000x g for 15 minutes and 25μl of DEPC-H_2O added to the Eppendorf tube.

2B.4.5 Reverse Transcription of RNA

Assay conditions
50mM Tris-HCl (pH 8.3)
75mM KCl
10mM DTT
3mM $MgCl_2$
0.5mM each of dGTP, dATP, dTTP, dCTP
10μg/ml oligonucleotide*

* oligonucleotide = $pd(T)_{15}$ or $pd(N)_6$

10μl of each RNA sample was incubated in the above conditions, in a total volume of 20μl, at 65°C for 3 minutes. 200 units of M-MLV Reverse Transcriptase were then added to each sample and the mixture incubated at 37°C for 1 hour. These two steps were repeated once more and followed by incubation at 70°C for 5 minutes. 2μl of each reaction mixture was then used directly in the Polymerase Chain Reaction as described below.

2B.5 POLYMERASE CHAIN REACTION

5x PCR Buffer	100mM Tris-HCl pH8.3
	500mM KCl
	15mM $MgCl_2$
	0.1% (w/v) gelatine
	0.1% (v/v) Tween 20
	0.1% (v/v) NP40

All reactions were carried out using either a Techne Programmable Dri-Block PHC1-1 or a Biometra Thermo-Trioblock.

To 100ρg-1ηg of DNA were added: 2μl of 10mM dNTPs, 10μl PCR Buffer, approximately 100ρmol of each oligonucleotide and 0.5μl of *Taq* Polymerase. H_2O was added to give a total volume of 50μl, and 50μl of mineral oil was layered on top of this to prevent evaporation of the reaction constituents. Annealing temperature and extension times were dependent on the nature of the DNA being amplified. Annealing temperature was determined by the following calculation:

annealing temperature (°C) = 4x (total G+C in oligonucleotide) +
2x (total A+T in oligonucleotide) - 5°C

Only those nucleotides which were complementary to the original DNA sequences were included in the above calculation and other nucleotides, for example those included in the oligonucleotide to create a restriction enzyme recognition site, were excluded.

Extension times used were : 1 minute for 1 to 400 bases
1-5 minutes for 400-1000bases.

Thus for each reaction:

denaturation:	93°C, 1 minute
annealing:	specific temperature, 1 minute
extension:	72°C, specific time

All reactions proceeded for 30 cycles.

In the case of samples derived from reverse transcription of RNA (2B.4.5), an additional 3 cycles were employed at the start. These were: 95°C for 5 minutes, specific temperature for 2 minutes and 72°C for 1 minute.

Positive controls utilised plasmid DNA to which primers were known to hybridise. Negative controls were identical to test samples but lacked DNA.

To remove oligonucleotides and dNTPs from the reaction products, reaction mixtures were electrophoresed through a 1% low melting point agarose gel and the DNA of the appropriate size cut out of the gel and eluted using either the hot phenol method or Geneclean II (Bio 101 Inc.).

2B.6 *E. coli* TRANSFORMATION

2B.6.1 Transformation of *E. coli* with plasmid DNA

TSB

10% (w/v) PEG 3000 (molecular weight approximately 3000)
5% DMSO (v/v)
10mM $MgCl_2$
10mM $MgSO_4$
Made to final volume with L-Broth

50 ml of L-Broth was innoculated with 0.5 ml of a stationary-phase culture of W3110Iq, TG1 or mutL cells and shaken at 37°C until OD_{650}=0.3. Cells were centrifuged at 4000x g for 5 minutes at room temperature and the cell pellet

resuspended in 0.1 volume of TSB, followed by incubation in glass tubes on ice for at least 30 minutes. DNA was added to 100μl of the cell suspension and this was incubated on ice for 30 minutes. 200μl TSB and 20μl 20% (w/v) glucose were added, and the mixture shaken at 37°C for 1 hour and then spread onto L-Amp agar plates. Plates were incubated for approximately 16 hours at 37°C.

2B.6.2 Transformation of *E. coli* with M13 Replicative Form or Single-Stranded DNA

The procedure followed was that for transformation of plasmid DNA up to and including the incubation of cells and DNA on ice. After this, 1ml of TSB and 20μl 20% (w/v) glucose were added and the mixture shaken at 37°C for 1 hour. 2.5ml molten (45°C) Top agar containing X-gal and IPTG was then added together with 200μl of *E. coli* cells in L-Broth of OD_{650}=0.3 and the mixture was poured onto minimal agar plates and allowed to solidify before incubation at 37°C for approximately 16 hours.

2B.7 DNA SEQUENCING

2B.7.1 Small-Scale M13 Template Preparation

M13 TE 10mM Tris
0.1mM EDTA
pH8.0

A culture of TGI cells in L-Broth of OD_{650}=0.2 was innoculated with a fresh M13 plaque and shaken for 4-5 hours at 37°C. The culture was centrifuged at 17000x g for 2 minutes at room temperature and the cell pellet discarded. To the supernatant was added 200μl of 20% (w/v) PEG 6000 in 2.5M NaCl, followed by incubation at room temperature for 30 minutes. This was then centrifuged at 17000x g for 5 minutes at room temperature, the supernatant discarded and the centrifugation repeated. All

remaining supernatant was removed using a drawn-out capillary tube, and 100μl M13 TE was added, together with 50μl phenol. This was vortexed vigorously for 10 seconds, incubated at room temperature for 10 minutes, vortexed for 10 seconds and centrifuged at 17000x g for 5 minutes at room temperature. The organic layer was discarded and the single-stranded DNA was recovered from the aqueous phase by ethanol precipitation. The DNA pellet was resuspended in 50μl of M13 TE.

2B.7.2 Dideoxynucleotide DNA Sequencing

TM	100mM Tris pH 8.5
	50mM $MgCl_2$
Chase	0.25mM dGTP
	0.25mM dATP
	0.25mM dCTP
	0.25mM dTTP

dNTP Termination Mixes (volumes in μl)

	G	A	T	C
0.5mM dGTP	6.25	125	125	125
0.5mM dTTP	125	125	6.25	125
0.5mM dCTP	125	125	125	6.25
10mM ddGTP	4			
10mM ddATP		0.25		
10mM ddTTP			12.5	
10mM ddCTP				2
M13 TE	250	125	250	250

Formamide Dye

0.1%(w/v) xylene cyanol
0.1%(w/v) bromophenol blue
10mM EDTA
in formamide

Primer Mix

8μl DNA template (approximately 1μg)
1μl oligonucleotide primer (approximately 0.3 ρmole for 17-mer)
1μl TM

Primer Mix for Single Nucleotide Tracking

3μl DNA template(approximately 0.4μg)
0.3μl oligonucleotide primer (approximately 0.1 ρmole for 17-mer)
0.3μl TM
2.4μl H_2O

Klenow Mix (per template)

8μl 10mM Tris pH8.0
0.8μl 100mM DTT
4μCi [α-^{35}S]dATP (8μCi/μl; Amersham)
1.7 units *E.coli* DNA Polymerase I Klenow Fragment (5U/μl)

Oligonucleotide Primers

705C: 5′-TCA GGA GAC TCT AAG GC-3′
complementary to nucleotides 121-136 of the sense strand of the C gene of HBV.

159E: 5′-GGC CCA CAT TAG TGT TG-3′

complementary to nucleotides 267-283 of the sense strand of the C gene of HBV.

193C: 5′-ACG CGG CGA TTG AGA TC-3′

complementary to nucleotides 500-516 of the sense strand of the C gene of HBV.

542E: 5′-ATT AGG AAA AGA GGG CG-3′

complementary to nucleotides 633-649 of HBV130 (Pasek *et al.*, 1989), 5′to the stop codon of the C gene and within the P ORF.

The method used was essentially that of Sanger *et al.* (1977). Template, normally made by the small-scale template method, was hybridised to an oligonucleotide complementary to a site within either the cloned insert or the *lacZ* gene of M13mp18, by incubation of the template/primer mix at 80°C for 5 minutes followed by cooling to room temperature over approximately 15 minutes. Sequencing reactions were then performed in microtitre plates as follows. To each well was added: 2μl of G, A, T or C mix, 2μl of template/primer mix and 2μl of Klenow mix. These were mixed at the bottom of each well and incubated at room temperature for 25 minutes. 2μl of Chase mix was then added to each well, mixed, and incubated at room temperature for 20 minutes. Reactions were stopped by addition of Formamide Dye (2μl) followed by incubation at 80°C for 15 minutes, then cooling on ice before loading onto a polyacrylamide gel.

2B.7.3 Single Nucleotide Tracking

This was used for rapid screening of DNA produced by site-directed mutagenesis and the method is essentially that for normal sequencing after the initial hybridisation, but only one nucleotide termination mix was used.

2B.7.4 Urea-Polyacrylamide Gel Electrophoresis of Sequencing Reaction Products

Urea-Polyacrylamide Gel Mix (for 6% gel)

17g urea
6ml 38%(w/v) Acrylamide, 2%(w/v) Bis-acrylamide
4ml 10x TBE
0.24ml 10%(w/v) ammonium persulphate
0.035ml TEMED (N,N,N,N'-tetraethylenediamine)
H_2O to final volume of 40 ml.

Gel mix was poured between 2 glass plates (40cm x 20cm) separated by spacers (0.5mm) and allowed to polymerise with the well comb in place.

Approximately 2μl of each sample was loaded onto the gel, in the order GATC, and the sequencing reaction products separated by electrophoresis in 1xTBE running buffer at a constant power of 25 watts. Electrophoresis was continued until the required separation was achieved (bromophenol blue migrates with DNA fragments of approximately 25 nucleotides in length, xylene cyanol migrates with DNA fragments of approximately 75 nucleotides in length). After electrophoresis, the gel was transferred to blotting paper and dried on a vacuum gel dryer. X-ray film was exposed to the gel for at least 16 hours in a light-excluding cassette, at -70°C to prevent the gel from sticking to the film.

2B.8 SITE-DIRECTED MUTAGENESIS

2B.8.1 Large-Scale Single-Stranded M13 Template Preparation

200μl of TG1 cells, grown overnight at 37°C in L-Broth, were added to 20 ml fresh L-Broth and shaken at 37°C for 3 hours. 100μl of this culture was then added to 1ml L-Broth, which was also innoculated with a fresh M13 plaque, and this was shaken

at 37°C for 4 hours. After centrifugation at 17000x g for 5 minutes at room temperature, the supernatant was removed and added to a 100ml L-Broth culture of TG1 cells (OD_{550} of 0.3). This was shaken at 37°C for 4 hours and then centrifuged at 5000x g for 30 minutes at 4°C. To the supernatant, 0.2 volume of 20% (w/v) PEG 6000 in 2.5M NaCl was added, mixed thoroughly and incubated for 1 hour at 4°C. The phage precipitate was pelleted by centrifugation at 15000x g for 20 minutes at room temperature and the supernatant removed. Re-centrifugation (15000x g for 5 minutes) was followed by removal of all remaining PEG/NaCl. The phage pellet was resuspended in 500μl M13 TE and centrifuged for 5 minutes at 17000x g at room temperature to remove any remaining cells. 200μl of 20% PEG (w/v) in 2.5M NaCl was mixed with the supernatant and incubated at room temperature for 15 minutes. All supernatant was removed after centrifugation at 17000x g for 5 minutes at room temperature and re-centrifugation at 17000x g at room temperature for 2 minutes. The viral pellet was then resuspended in 200μl phenol, vortexed for 15 seconds, incubated at room temperature for 15 minutes, vortexed for 15 seconds and centrifuged for 3 minutes at 17000x g. The aqueous layer was removed and subjected to phenol extraction as before. 500μl diethyl ether was then mixed thoroughly with the supernatant and the top layer discarded. This was repeated 3 times. The aqueous phase was then mixed with 500μl chloroform and the organic layer discarded. This was repeated once. The single-stranded M13 DNA was then precipitated with ethanol, dried and suspended in M13 TE to a final concentration of 1μg/μl.

2B.8.2 Phosphorylation of 5′ ends of oligonucleotides with T4 polynucleotide kinase

10X LK Buffer

0.5M Tris pH7.5
0.1M $MgCl_2$
0.1M DTT
10mM ATP
0.48mg/ml BSA

5′ phosphate groups were added to unphosphorylated oligonucleotides by polynucleotide kinase. 1 unit of T4 polynucleotide kinase was incubated with approximately 100pmole of oligonucleotide in 1x LK Buffer for 90 minutes at 37°C. The reaction was stopped by incubation at 70°C for 10 minutes. After PCI extraction, the oligonucleotide was precipitated with ethanol and resuspended in 12μl of H_2O.

2B.8.3 Mutagenesis

To 12μl of phosphorylated oligonucleotide were added 3μg of template and approximately 100pmole of M13 universal primer (-20) (New England Biolabs). This was incubated in a final concentration of 1xTM at 80°C for 5 minutes and then cooled to room temperature, allowing the oligonucleotides to anneal to the template. To this mixture was then added 1.5 units of DNA Polymerase I (Klenow fragment) and 5 units of T4 DNA ligase. This was incubated in the presence of 100μM dNTPs, 1mM ATP, 0.5mM Tris pH8.0 and 5μM EDTA, for 90 minutes at room temperature followed by one extraction with phenol and one extraction with butanol. The DNA was then precipitated with ethanol, resuspended in 32μl of H_2O and digested with *EcoR*I and *BamH*I. The digestion products were run in a 1% pure low melting-point agarose gel and the appropriate band of DNA cut out and eluted by the warm phenol method.

250ηg of this eluted fragment was ligated to 80ηg of M13mp18 DNA which had previously been digested with *EcoR*I and *BamH*I, precipitated with isopropanol to remove the small fragment generated by the digestion, and resuspended in an appropriate volume of H_2O. Products of this ligation were transformed into mutL cells and plated onto minimal agar plates containing X-gal and IPTG. White plaques were picked, small-scale template preparations made and the appropriate nucleotide screening performed, followed by complete sequencing of positive templates.

2B.9 PROTEIN METHODS

2B.9.1 Preparation of Proteins from *E. coli*

Lysis Buffer 50mM Tris-HCl pH 8.0
1% (w/v) Triton X-100

5ml of a stationary-phase culture of the appropriate plasmid-carrying cells, grown in selecting medium, were added to 500ml of L-Broth containing ampicillin (pre-warmed to 37°C) and shaken at 37°C until OD_{650} = 1.0. IPTG was then added to a final concentration of 0.5mM and shaking at 37°C continued for a further 16 hours. Cells were harvested by centrifugation at 10000x g for 10 minutes at 4°C and the cell pellet resuspended in 5ml of Lysis Buffer. This suspension was split between eppendorf tubes and each tube subjected to sonication (3 x 15 seconds). Cell debris was removed by centrifugation at 11000x g for 10 minutes at 4°C and the supernatant (the crude extract) retained.

2B.9.2 Further Purification of Crude Extract

Column Buffer 0.1M $NaHCO_3$ pH 7.0
0.1mM Phenylmethylsulphonamide (PMSF)
0.1mM Benzamidine

Crude extract underwent 35% ammonium sulphate precipitation by addition of 2g of ammonium sulphate per 10ml of extract, gradually over 30 minutes on ice with constant stirring, followed by a further 30 minutes of stirring on ice. The precipitate was recovered by centrifugation at 13000x g for 10 minutes at 4°C. The pellet was then resuspended in 50mM Tris pH 8.0 and dialysed against 50mM Tris pH 8.0 for 16 hours at 4°C.

This was then applied to a 3cm x 10cm, or a 3cm x 40cm Sepharose 4B column equilibrated with Column Buffer. 5ml fractions were collected and 20μl or 50μl of each fraction electrophoresed through a 15% polyacrylamide gel. Those fractions containing the most, and the most pure, desired proteins were then pooled and subjected to centrifugation at 100000x g for 1 hour at 4°C and the pellet resuspended in 50mM Tris pH 8.0.

2B.9.3 Electrophoresis of Proteins in Polyacrylamide Gels

Resolving gel Solution (sufficient for 4 gels)

8.8ml 30% (w/v) acrylamide, 0.2% (w/v) bis-acrylamide
0.35ml 10% (w/v) SDS
7.0ml 1.875M Tris-HCl pH 8.8
10.09ml H_2O
0.05ml 10% (w/v) ammonium persulphate (fresh)
0.02ml TEMED

Stacking Gel Solution

1.67ml 30% (w/v) acrylamide, 0.2% (w/v) bis-acrylamide
1.0ml 0.6M Tris-HCl pH 6.8
0.1ml 10% (w/v) SDS
7.18ml H_2O
0.05ml 10% (w/v) ammonium persulphate (fresh)
0.01ml TEMED

Sample Loading Buffer

2ml 0.6M Tris-HCl pH 6.8
2ml glycerol
0.4g SDS
5.6ml H_2O
0.01g Bromophenol Blue

10 x Electrophoresis Buffer

25mM Tris-HCl
192mM Glycine
0.1% (w/v) SDS

Stain

0% (v/v) methanol
10% (v/v) glacial acetic acid
0.2% (w/v) Coomassie Blue

De-Stain

40% (v/v) methanol
10% (v/v) glacial acetic acid

All electrophoresis was carried out using the Biorad MiniProtean II Gel Apparatus and the method used was the discontinuous buffer system of Laemmli (1970).

Resolving gel solution was poured between 2 glass plates (10cm x 7cm) separated by 0.75mm spacers and left to polymerise at room temperature for at least 1 hour after being overlaid with H_2O. The H_2O was then removed and Stacking Gel Solution added. The gel comb was inserted and the gel left to polymerise at room temperature for at least 20 minutes.

Samples were loaded and electrophoresis carried out in 1 X Electrophoresis Buffer at 50 volts until the dye had entered the resolving gel, and then at 200 volts until the dye reached the bottom of the gel. The gel was then removed from the apparatus and either transferred to nitrocellulose (see below) or stained as follows. The gel was immersed in Stain and rotated gently for at least 20 minutes. The stain was removed, De-Stain added and rotation continued until the areas of the gel not containing protein were colourless.

Molecular weight markers used were:

Low Molecular Weight Markers (Pharmacia)

94kD	phosphorylase b
67kD	bovine serum albumin
43kD	ovalbumin
30kD	carbonic anhydrase
20.1kD	trypsin inhibitor
14.4kD	α-lactalbumin

Rainbow Molecular Weight Markers (Amersham)

46kD	ovalbumin
30kD	carbonic anhydrase
21.5kD	trypsin inhibitor
14.3kD	lysozyme
6.5kD	aprotinin
3.4kD	insulin (b) chain
2.35kD	insulin (a) chain

2B.9.4 Preparation of Protein Samples for Electrophoresis

A. Reducing Electrophoresis

An equal volume of Sample Loading Buffer was added to each sample, together with DTT to a final concentration of approximately 0.1M. This was then incubated at 100°C for 5 minutes (except where stated otherwise) and loaded immediately onto the gel.

B. Non-Reducing Electrophoresis

As for Reducing Electrophoresis samples, with the omission of DTT.

2B.9.5 Lowry Protein Assay

Solution 1	2% (w/v) Na_2CO_3
	0.1M NaOH
	0.01% (w/v) $CuSO_4.5H_2O$
	0.02% (w/v) sodium citrate

Solution 2	50% Folin-Cocteau Reagent
	50% H_2O

0.4 ml of sample was added to 2ml of Solution 1, mixed and incubated at room temperature for 10 minutes. 0.2ml of Solution 2 was then added, mixed thoroughly and incubated at room temperature for 30 minutes. Optical density at 550ηm was then measured. The assay was calibrated using BSA standards of 1-200μg/ml.

2B.9.6 Electrophoretic Transfer of Proteins Separated in Polyacrylamide Gels to Nitrocellulose Membranes (Western Blotting)

Transfer Buffer	25mM Tris-HCl pH8.3
	192mM Glycine
	20% (v/v) Methanol
Amido Black Stain	0.1% (w/v) amido black NB (Sigma)
	40% (v/v) Methanol
	10% (v/v) Glacial Acetic Acid
Amido Black de-Stain	40% (v/v) Methanol
	10% (v/v) Glacial Acteic Acid

The method used was essentially that described by Towbin *et al.* (1979). Samples were subjected to SDS-PAGE in a 15% polyacrylamide gel which was then laid onto a piece of nitrocellulose of equal size. These were then placed between 6 pieces of blotting paper cut to the size of the gel, and 2 Scotchbrite pads, all of which had been soaked in Transfer Buffer. This sandwich was then inserted into a "Trans-blot" apparatus (Biorad) filled with Transfer Buffer. Proteins were transferred to the nitrocellulose at 60 V for 1 hour. The temperature was kept low throughout by the presence of a water-filled cooling coil.

After transfer, nitrocellulose containing the marker lanes was cut off and stained irreversibly with amido black for 30 seconds and then washed in several changes of Amido Black de-stain. This step was not required when pre-stained markers (Rainbow Molecular Weight Markers) were used.

2B.9.7 Isolation of anti-HBcAg IgG from Rabbit Serum (for use in Western Blotting)

PBS	4mM NaH_2PO_4
	6mM Na_2HPO_4
	150mM NaCl
	pH adjusted to 7.0

Serum from the third bleed withdrawn 9 days after the third innoculation of rabbit 87, with HBcAg produced in *E. coli* (Biogen Inc.), was diluted with an equal volume of PBS, followed by addition of ammonium sulphate, gradually over 30 minutes, to 40% saturation followed by stirring on ice for a further 30 minutes. This was centrifuged at 13000x g for 10 minutes at 4°C and the pellet resuspended in PBS. The 40% ammonium sulphate precipitation was then repeated and the pellet again resuspended in PBS and dialysed for at least 16 hours against PBS at 4°C.

2B.9.8 Immunological Detection of Proteins Immobilised on Nitrocellulose Filters

TS	10mM Tris-HCl pH 8.1
	150mM NaCl
Block Buffer	5% BSA in TS

Developing Solution	100mM Tris-HCl pH 9.5
	100mM NaCl
	50mM $MgCl_2$
	0.33mg/ml nitro-blue-tetrazolium (NBT)
	0.25mg/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP)
	(NBT and BCIP added immediately prior to use)

The nitrocellulose membrane onto which protein had been electroblotted was incubated for at least 1 hour at room temperature, with shaking, in Block Buffer in order to prevent non-specific binding of antibody to the membrane. Block Buffer was removed and the membrane incubated with the primary antibody (anti-HBcAg IgG at approximately 5μg/ml in Block Buffer) for 4-16 hours at room temperature with gentle shaking. The primary antibody solution was removed and the nitrocellulose washed in 5 changes of TS over 30 minutes. Anti-rabbit alkaline phosphatase conjugate (5μg/ml in Block Buffer) was then added and incubated for 2 hours, gently shaking at room temperature. The nitrocellulose was then washed in TS as before and developed with Developing Solution for 2-15 minutes at room temperature. The reaction was stopped by washing the membrane in several changes of H_2O.

2B.10 RADIOIMMUNE ASSAYS

2B.10.1 Isolation of anti-HBcAg/anti-HBeAg from Human Serum

This was performed by the Department of Medical Microbiology of the University of Edinburgh.

Solution A: 0.02M NaH_2PO_4

Solution B: 0.02M Na_2HPO_4

0.01M Phosphate Buffer pH 7.5 = 190ml Solution A + 1000ml Solution B

2.5ml of high titre human anti-HBcAg/anti-HBeAg serum was UV-treated for 30 minutes and then inactivated at 56°C for 30 minutes. This was then centrifuged at 3000 rpm for 5 minutes and the supernatant placed in dialysis tubing and dialysed for approximately 16 hours at 4°C against 0.01M Phosphate Buffer. 12.5g of DEAE cellulose (Whatman DE52) was equilibrated with Phosphate Buffer and to it was added the inactivated serum which was then mixed gently for 10 minutes with a metal applicator, followed by incubation at 4°C for 30 minutes with occasional stirring. 50ml of Phosphate Buffer was added to the cellulose/serum mixture and filtered through a sintered glass funnel, a further 50ml Phosphate Buffer added and all filtrate retained. 10ml 1.5M NaCl was added to 100ml of filtrate and concentrated to the original volume of serum by using dialysis tubing under vacuum.

2B.10.2 ^{125}I-Labelling of anti-HBcAg/anti-HBeAg Isolated from Human Serum

This was performed by Peter McCulloch of the Department of Medical Microbiology of the University of Edinburgh.

The method used was essentially that of Hunter and Greenwood (1962). 10μg of anti-HBcAg/HbeAg (2B.10.1) was labelled with 500μCi ^{125}I and diluted in RIP Buffer (2B.10.3) to give a final concentration of 50000 cpm per 150μl.

2B.10.3 Microtitre Plate-Based Radioimmune Assay

Coating Solution

120μg/ml anti-HBeAg IgG
60μg/ml anti-HBcAg IgG

RIP Buffer 0.5% BSA in PBS

100μl of Coating Solution was added to a microtitre plate well and incubated at room temperature for 15-30 minutes. The Coating Solution was then removed and the plate

washed 3 times with H_2O and allowed to dry at room temperature. 75μl of sample (diluted in RIP Buffer) was added to each well and incubation at room temperature continued for 16 hours. The sample was then removed and the plate washed 10 times with H_2O and air-dried at room temperature. ^{125}I -labelled anti-HBcAg IgG (1000cps) was added to each well and incubated for 1 hour at 45°C. This was then removed, the wells washed 10 times with H_2O and the ^{125}I-IgG remaining bound counted with a gamma counter (LKB 1275 MiniGamma).

2B.10.4 Bead-Based Radioimmune Assay

Beads (diameter 6.4mm) were incubated in 200μl Coating Solution per bead for 15-30 minutes at room temperature, then Coating Solution was removed and the beads air-dried at room temperature. 150μl of sample (diluted in RIP Buffer) was added to each bead and incubation proceeded for 16 hours at room temperature. Sample was removed and the beads washed 10 times with H_2O, after which ^{125}I-labelled polyclonal anti-HBcAg/anti-HBeAg IgG (1000cps) was added to each bead and incubated for one hour at 45°C. After removal of ^{125}I-IgG, beads were washed 10 times with H_2O and the ^{125}I-IgG remaining bound was counted with a gamma counter (LKB 1275 MiniGamma) for 10 minutes.

2B.10.5 Correction Factor

As each sample was counted for 10 minutes, between the counting of the first and last samples, significant ^{125}I decay occurred. In order that all samples could be compared directly, a correction factor was employed to remove the effect of ^{125}I decay, and was calculated as follows:

^{125}I half-life = 60 days = 86400 minutes
i.e. after 86400 minutes ^{125}I activity decays by a factor of 0.5, so in time t, the decay factor is $0.5t/86400$.

Therefore the amount of decay (in cpm) of the original activity, x (in cpm) $= x \times (0.5t/86400)$, which in turn $= x - y$, where y is the actual reading in cpm recorded by the gamma counter. i.e. $x \times (0.5t/86400) = x - y$

which can be re-arranged to give: $1/x = 1/y - t/172800y$.

As the values for t and y are known, the value for x (the activity before decay) can be calculated.

CHAPTER 3: The Amino-Terminal Region of HBcAg is Important for Core Particle Structure Determination

3.1 Introduction

The amino-terminal region of HBcAg is highly conserved among mammalian hepadnaviruses. Indeed, amino acids 1 to 10 are completely conserved (Figure 3.1A) and 26 of the corresponding nucleotides (1 to 30) are identical (Figure 3.1B). This would suggest an important role for these residues. Indeed, Salfeld *et al.* (1989) found that a protein synthesised *in vitro* which consisted of amino acids 12-183 of HBcAg had drastically reduced activity in a radioimmune assay with anti-HBcAg antibodies, compared to wild-type HBcAg.

Further evidence for the importance of the amino-terminal region came from Stahl *et al.* (1982). With the aim of producing a plasmid from which HBcAg could be expressed in *E. coli* with enhanced efficiency, several plasmid constructs were created (Figure 3.2A). Ammonium sulphate precipitated extracts of cultures of *E. coli* containing these plasmids were investigated in a radioimmune assay with polyclonal anti-HBcAg antibodies (as in section 2B.10.3) (Figure 3.2B). A striking difference was apparent between the activity of the pRI-11 and pRI-4 products. The proteins produced from these plasmids differ in only two amino acids. Both plasmids encode amino acids 1 to 8 of β-galactosidase followed by three linker amino acids, but pRI-11 then encodes amino acids 3 to 183 of HBcAg while pRI-4 lacks HBcAg amino acid codons 3 and 4 and encodes amino acids 5 to 183. The product of pRI-11 was shown to form particles morphologically and immunologically indistinguishable from HBV nucleocapsids isolated from infected individuals, and has been subsequently used extensively in studies of HBcAg.

The product of pRI-4 has been investigated further in this work, in order to determine whether its lack of activity in the radioimmune assay was due to the level of expression of the protein in *E. coli*, which was not previously investigated, or to a difference in the structure of the core particles it forms, if indeed it forms core particles, thus reducing its affinity for anti-HBcAg antibodies. Similar investigations were also undertaken of a panel of substitution and deletion mutants involving amino

Legend 3.1

A. Conservation of amino acids 1 to 10 between HBV subtypes and the other mammalian hepadnaviruses.

B. Conservation of nucleotides 1 to 30 between HBV subtypes and the other mammalian hepadnaviruses.

A

	1 10
HBV*adyw*	MDIDPYKEFG
HBV*adr*	MDIDPYKEFG
HBV*adw*	MDIDPYKEFG
HBV*ayw*	MDIDPYKEFG
HBV*ayr*	MDIDPYKEFG
HBV*adw2*	MDIDPYKEFG
HBV*adr4*	MDIDPYKEFG
WHV	MDIDPYKEFG
GSHV	MDIDPYKEFG
IDENTITY	MDIDPYKEFG

B

	1 30
HBV*adyw*	ATGGACATTG ACCCTTATAA AGAATTTGGA
HBV*adr*	ATGGACATTG ACCCGTATAA AGAATTTGGA
HBV*adw*	ATGGACATTG ACCCTTATAA AGAATTTGGA
HBV*ayw*	ATGGACATCG ACCCTTATAA AGAATTTGGA
HBV*ayr*	ATGGACATTG ACCCGTATAA AGAATTTGGA
HBV*adw2*	ATGGACATTG ACCCTTATAA AGAATTTGGA
HBV*adr4*	ATGGACATTG ACCCGTATAA AGAATTTGGA
WHV	ATGGACATAG ATCCTTATAA AGAATTTGGT
GSHV	ATGGACATAG ATCCCTATAA AGAATTTGGT
IDENTITY	ATGGACAT_G A_CC_TATAA AGAATTTGG_

Figure 3.2

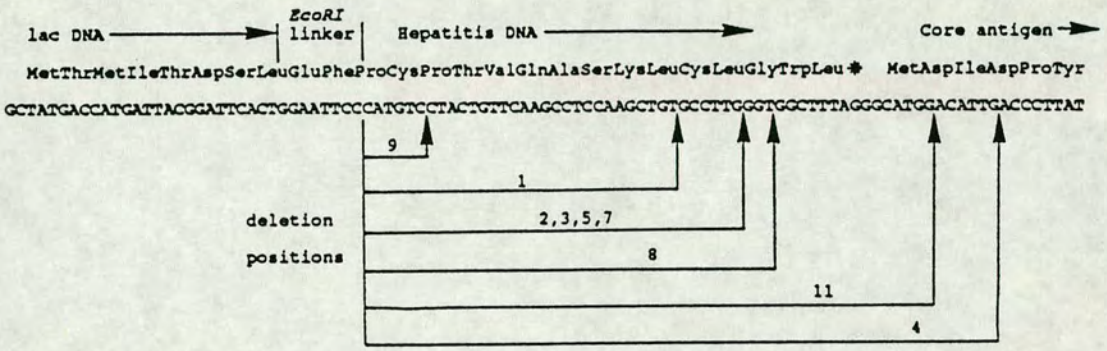
(from Stahl *et al.*, 1982)

A. Sequences of plasmid constructions produced by Stahl *et al.* (1982). The region shown extends from the start point of translation of the β-galactosidase gene to amino acid 6 of the HBcAg gene. The deletions extend from nucleotide 2 of the Pro codon following the EcoRI linker, to the points indicated by the arrow heads. Note particularly deletions 4 and 11.

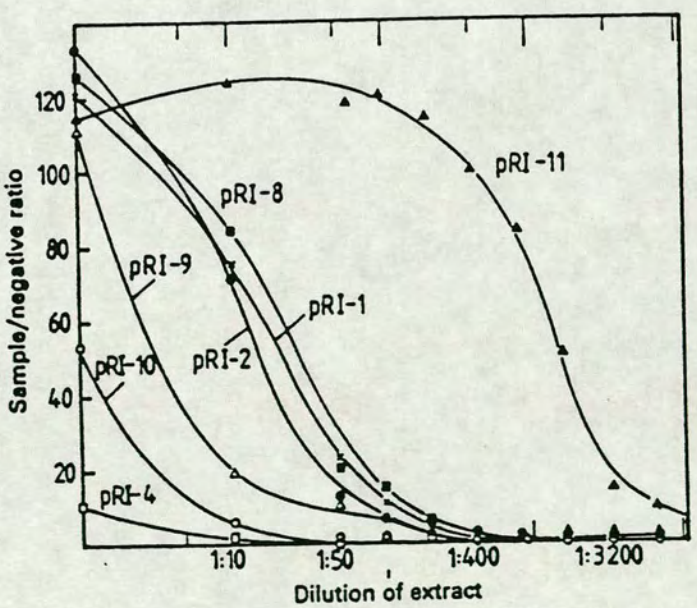
* = stop codon.

B. Results of radioimmune assay of ammonium sulphate-precipitated extracts of *E. coli* cultures expressing the plasmids described in A (above), with polyclonal anti-HBc/HBe antibodies (Stahl *et al.*, 1982). Note particularly the behaviour of pRI-4 and pRI-11.

A


lac DNA
EcoRI linker
Hepatitis DNA
Core antigen
MetThrMetIleThrAspSerLeuGluPheProCysProThrValGlnAlaSerLysLeuCysLeuGlyTrpLeu*
MetAspIleAspProTyr
GCTATGACCATGATTACGGATTCACTGGAATTCCCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTAGGGCATGGACATTGACCCTTAT
9
1
deletion
positions
2,3,5,7
8
11
4

B


Sample/negative ratio
120
100
80
60
40
20
pRI-11
pRI-8
pRI-9
pRI-1
pRI-2
pRI-10
pRI-4
1:10
1:50
1:400
1:3200
Dilution of extract

acids 3 to 6, which were created in this work by site-directed mutagenesis.

3.2 Results

3.2.1 Plasmid Constructions

Plasmids pRI-11 and pRI-4 contain HBcAg sequences under control of the *lacUV5* promoter (section 2A.2.2). The *tac* promoter had been shown to be more efficient for HBcAg expression in *E. coli* (Stahl and Murray, 1989), and so the coding regions of pRI-11 and pRI-4 were transferred to a vector in which they could be expressed under control of the *tac* promoter in order to maximise their expression.

Plasmid pTacpCore contains sequences encoding amino acids 1-8 of β-galactosidase fused to 3 linker amino acids (Glu Phe His) which are in turn fused to amino acids 3-183 of HBcAg. These sequences are expressed under control of the *tac* promoter, expression from which is induced by the presence of IPTG. This plasmid was created by the digestion of pRI-11 (section 2A.2.2) with *EcoR*I and *BamH*I to liberate a 1005bp fragment containing the coding sequences described above, minus the β-galactosidase amino acid codons; which was then ligated to the larger (4373bp) *EcoR*I/*BamH*I digestion product of pTacHpaIIR2 (section 2A.2.2), which contains the *tac* promoter and the coding sequences for amino acids 1 to 8 of β-galactosidase. The resulting plasmid was used to express large amounts of HBcAg in *E. coli*, referred to in the remainder of this thesis as wtHBcAg or as CWT.

Plasmid pTacpRI-4 was created in the same manner as pTacpCore but with pRI-4 (section 2A.2.2) substituted for pRI-11, thereby creating a final product which encodes amino acids 1-8 of β-galactosidase fused to three linker amino acids (Glu Phe His) which are in turn fused to amino acids 5-183 of HBcAg.

The sequences of pTacpCore and pTacpRI-4 were confirmed by DNA sequencing (Figure 3.3).

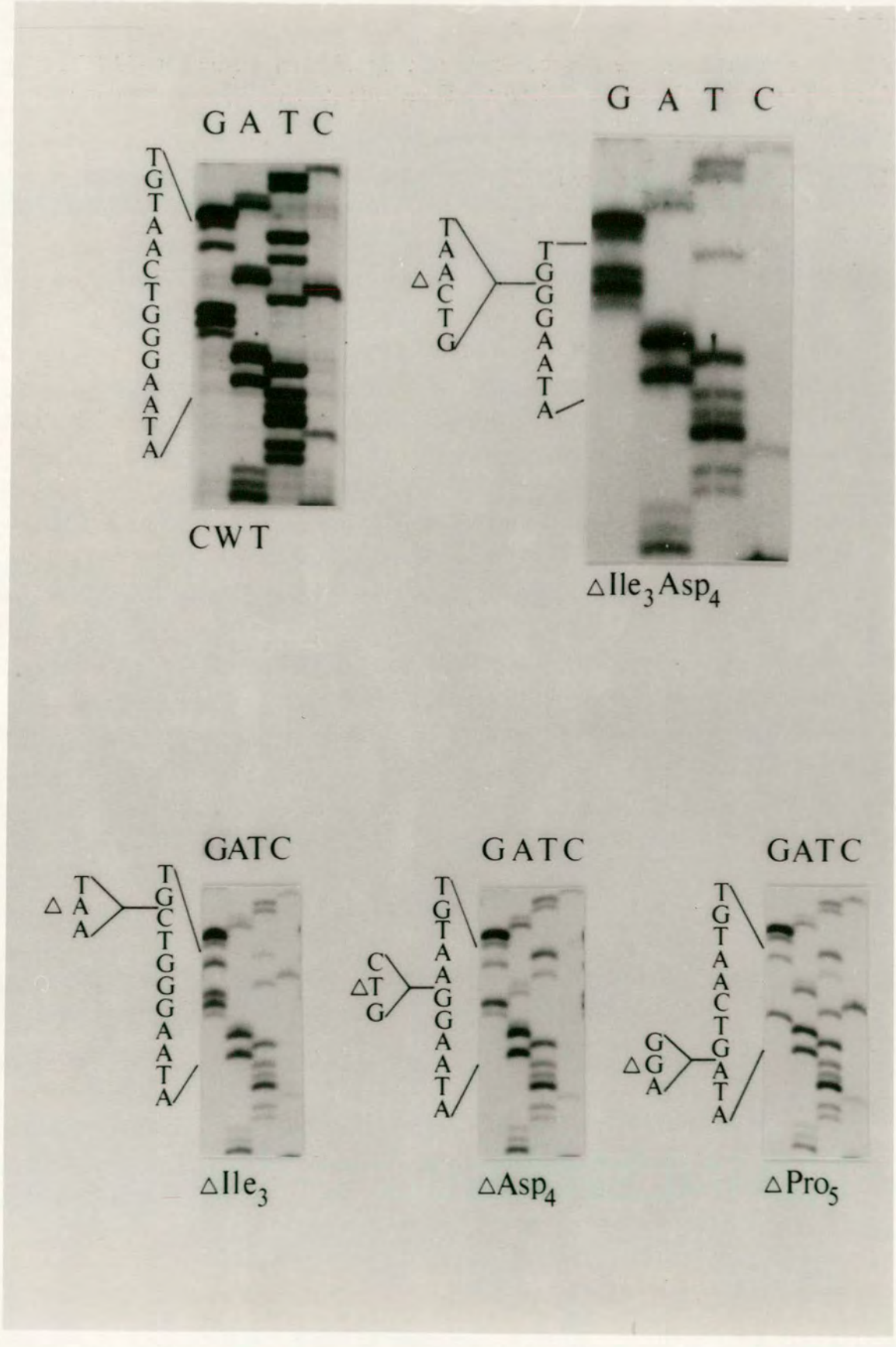
Figure 3.3

Wild-type and deletion mutants of the coding sequences for the amino-terminal region of HBcAg

All sequences shown are the non-coding strand and read, from left to right, GATC and from top to bottom, 3′ to 5′. All show the EcoRI site (GAATTC) at the top of each photograph.

The deleted codons are indicated by Δ.

CWT is pTacpCore
ΔIle_3Asp_4 is pTacpRI-4.


G A T C
TGTAACTGGGAATA
CWT
G A T C
△
TAACTG
TGGGAATA
△Ile3Asp4
GATC
△
TAA
TGCTGGGAATA
△Ile3
GATC
C
△TG
TGTAAGGAATA
△Asp4
GATC
GGA
△
TGTAACTGATA
△Pro5

3.2.2 Site-Directed Mutagenesis

The template for mutagenesis was created by insertion of the 1005bp fragment produced by *EcoRI/BamHI* digestion of pRI-11, between the *EcoRI* and *BamHI* sites of M13mp18, such that single-stranded template contained the "sense" strand. A large scale template preparation (2B.8.1) of this was prepared and used in site-directed mutagenesis.

Deletion Mutants

Three single deletion mutants were created using the following oligonucleotides:

ΔIle_3

oligonucleotide: 367E 5′ - CTT TAT AAG GGT CGT GGA ATT CGT AA - 3′

ΔAsp_4

oligonucleotide: 366E 5′ - ATT CTT TAT AAG GAA TGT GGA ATT CG - 3′

ΔPro_5

oligonucleotide: 365E 5′ - CCA AAT TCT TTA TAG TCA ATG TGG AAT T - 3′

These mutants were sequenced along the length of the coding region and are shown in Figure 3.3.

Five substitution mutations were also created, using the following oligonucleotides. The partners of the nucleotides to be mutated are underlined:

Ile_3 → Asp

oligonucleotide: 561C 5′ - TTA TAA GG TCA TCG TGG AAT TCG T - 3′

Asp_4 → Asn

oligoncleotide: 562C 5′ - TTA TAA GGG TTA ATG TGG A - 3′

Asp_4 → Val

oligonucleotide: 563C 5′ - TCT TTA TAA GGA ACA ATG TGG AAT T - 3′

Pro_5 → Gly

oligonucleotide: 564C 5′ - AAA TTC TTT ATA ACC GTC AAT GTC GAA - 3′

Tyr_6 → Phe

oligonucleotide: 565C 5′ - CCA AAT TCT TTA AAA GGG TCA ATG T - 3′

The sequences of these mutants were verified by DNA sequencing along the length of the coding region and are shown in Figure 3.4.

3.2.3 Expression and Purification

All plasmids were introduced into *E. coli* strain W3110Iq and expression of the desired proteins induced by exposure to 0.5mM IPTG. Crude extracts were made which were then subjected to ammonium sulphate precipitation at 35% ammonium sulphate saturation. Passage of dialysed ammonium sulphate precipitated extracts through a Sepharose 4B column resulted in all HBcAg-derived proteins present being eluted from the column before fraction 11. This corresponds to a species of very high

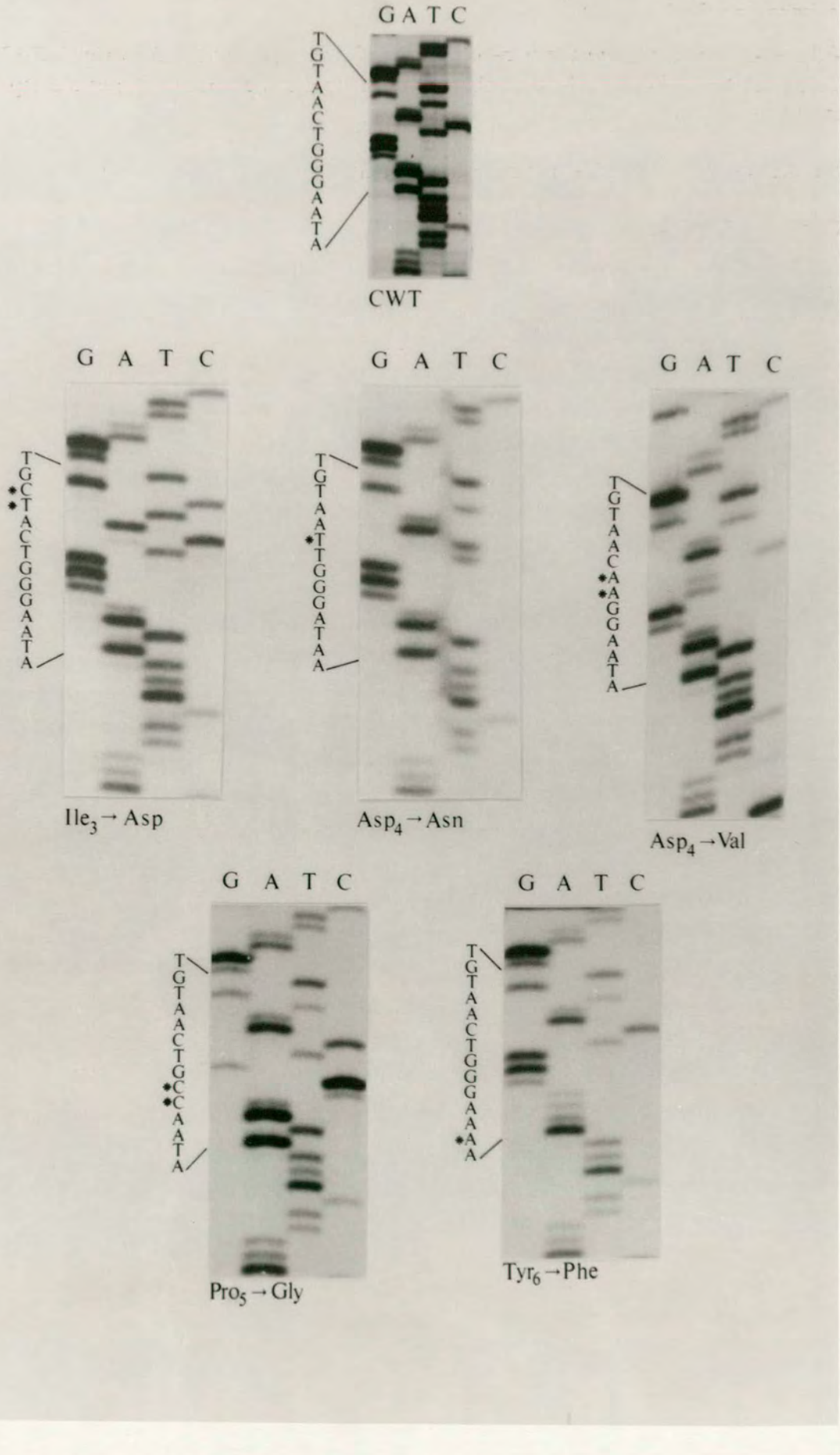
Figure 3.4

Wild-type and substitution mutants of the coding sequences for the amino-terminal region of HBcAg

All sequences shown are the non-coding strand and read, from left to right, GATC and, from top to bottom, 3′ to 5′. All show the EcoRI site (GAATTC) at the top of each photograph.

* indicates the nucleotides that have been mutated.

CWT is pTacpCore


G A T C
TGTAACTGGGAATA
CWT
G A T C
TGCTACTGGGAATA
Ile3 → Asp
G A T C
TGTAATTGGGATAA
Asp4 → Asn
G A T C
TGTAACAAGGAATA
Asp4 → Val
G A T C
TGTAACTGCCAATA
Pro5 → Gly
G A T C
TGTAACTGGGAAAA
Tyr6 → Phe

molecular weight, such as a core particle composed of 180 molecules of full-length protein, which has a predicted molecular weight of approximately 3.7 x 10^6D. The behaviour of all the mutant proteins during purification was therefore consistent with their existence in the form of high molecular weight particles.

A measurement of the levels of expression of the proteins was achieved by comparison of the final yield of each protein after purification. That the differences shown below were due to the amount of protein produced and not due to differential loss of protein during purification was verified by examination by eye of the amount of the desired protein present in reducing SDS-polyacrylamide gels of the crude extract, before purification (not shown).

Average yield from two 500ml cultures:

pTacRI-4 (ΔIle_3Asp_4)	50μg
ΔIle_3	75μg
$Asp_4 \rightarrow$ Asn	100μg
$Asp_4 \rightarrow$ Val	80μg
$Pro_5 \rightarrow$ Gly	750μg
$Tyr_6 \rightarrow$ Phe	850μg
pTacpCore	875μg

ΔAsp_4, ΔPro_5 and $Ile_3 \rightarrow$ Asp proteins could not be detected, either on Coomassie stained gels or on Western Blots. This was true when expression was attempted both in W3110Iq cells and in TG1 cells (not shown).

3.2.4 Detection of Messenger RNA

In order to determine whether the genes encoding the above proteins were being expressed, the presence or absence of C gene-specific mRNA within these cells was determined. This was achieved by making use of the polymerase chain reaction. Total RNA preparations were made from *E. coli* cultures containing the appropriate plasid, expression from which had been induced by IPTG (strains W3110Iq and TG1) as described in Section 2B.4.4. Half of this preparation was stored at -70°C while the remainder was treated with DNase (Section 2B.4.3) to remove any contaminating DNA. Finally, half of this DNase-treated RNA was further treated with RNase (Section 2B.4.4). All three RNA preparations were then subjected to two rounds of reverse transcription, using either the oligo dT primer or the random hexamer primer, and a proportion of each completed reaction mixture employed as template for the polymerase chain reaction. The following C gene-specific oligonucleotide primers were used in the amplification reaction:

140R: 5′-GGA ATT CCA CAT TGA CCC TTA-3′
which hybridises to the non-coding strand, beginning at the first nucleotide of the *EcoR*I site.

542E: 5′-ATT AGG AAA AGA GGG CG-3′
which hybridises to the coding strand 80 nucleotides downstream of the C gene stop codon.

and the annealing temperature used was 45°C.

All sequences between and including these oligonucleotide primers would be included in mRNA expressed from the *tac* promoter in these plasmids and the polymerase chain reaction would be expected to produce a DNA fragment of 630 bp, identified by electrophoresis of 5μl or 20μl of the PCR reaction mixture through a 1% agarose gel. The results of this procedure are summarised in Figure 3.5. As the PCR product (and

Figure 3.5

Indirect detection of C gene-specific mRNA by use of the polymerase chain reaction

RNA was isolated from *E. coli* in which expression of the appropriate plasmids had been induced. These samples were either un-treated, treated with DNase only or treated with DNase and RNase before being subjected to reverse transcription. An equal volume of the products of reverse transcription of each sample was then used as template for the polymerase chain reaction (PCR) using HBV C gene-specific primers. PCR reaction products were then electrophoresed through a 1% agarose gel and the presence or absence of a 630bp fragment detected by ethidium bromide staining. The results are summarised in this table.

✓ indicates the presence of the 630bp product

× indicates the absence of the 630bp product

The results produced were independent of the *E. coli* strain (W3110Iq or TG1) or reverse transcriptase primer (oligo dT or random hexamer) used and so each ✓ or × is representative of four separate results (with the exception of W3110Iq only and TG1 only, which are representative of two separate results).

Normal positive and negative PCR controls were used.

SAMPLE	RNA	RNA (DNase-TREATED)	RNA (DNase + RNase-TREATED)
RI-4 (ΔIle_3Asp_4)	✓	✓	×
ΔIle_3	✓	✓	×
ΔAsp_4	✓	✓	×
ΔPro_5	✓	✓	×
$Ile_3 \rightarrow Asp$	✓	✓	×
$Asp_4 \rightarrow Asn$	✓	✓	×
$Asp_4 \rightarrow Val$	✓	✓	×
$Pro_5 \rightarrow Gly$	✓	✓	×
$Tyr_6 \rightarrow Phe$	✓	✓	×
TacpCore	✓	✓	×
W3110Iq only	×	×	×
TG1 only	×	×	×

hence the reverse transcription product) was obtained when any contaminating DNA had been removed but not when RNA was removed, the PCR product was amplified from reverse-transcribed RNA, and as only a single band of the expected size was produced, the RNA was C gene-specific. Therefore, the lack of protein detected was not due to lack of mRNA expression or stability.

3.2.5 Electron Microscopy

Examination of the product of pTacpRI-4 (i.e. ΔIle_3Asp_4) with the electron microscope demonstrated that the particles it forms are morphologically indistinguishable form those formed from wild-type HBcAg (Figure 3.6). ΔIle_3Asp_4 has the largest mutation of the series examined (deletion of two amino acids) and reacts most abberantly with antibodies (see below), and so it may be predicted that the particles formed by the other mutant proteins also have morphology similar to that of wild-type particles. Therefore, the mutations made do not affect the ability of the proteins to form nucleocapsids-like structures.

3.2.6 Non-Reducing SDS-Polyacrylamide Gel Electrophoresis

Following incubation at 100°C for 5 minutes in the absence of reducing agent, fresh preparations of all proteins failed to enter a 15% SDS-polyacrylamide gel, or a small proportion of the sample ran as a faint smear (Figure 3.7), in the same manner as fresh preparations of CWT (see Chapter 4). In order for the protein to be prevented from entering the gel, it must have a molecular weight of greater than approximately 200kD. This would correspond to all protein molecules being present in the form of disulphide bond-linked multimers consisting of at least 10 monomers. This would require at least 80% of the potential disulphide bonds (see Chapter 4) which form the completely disulphide-linked network in the wild-type protein to be present.

Figure 3.6

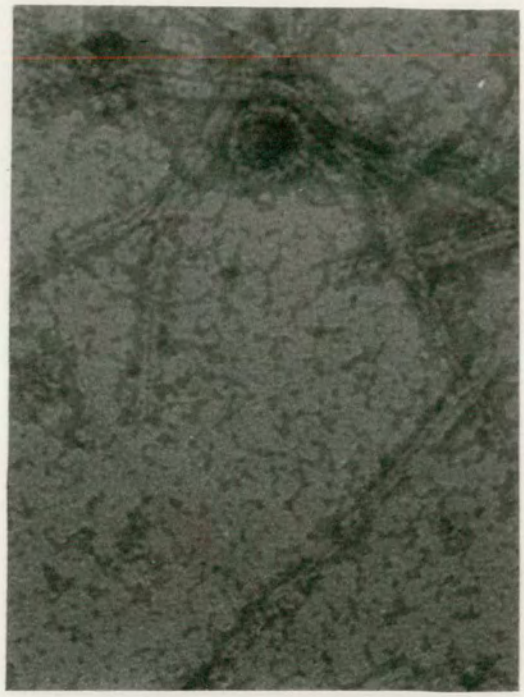
Electron Microscopy of TacpRI-4 (ΔIle_3Asp_4)

The purified protein product of pTacpRI-4 was examined with the electron microscope by S. Bury.

Both ΔIle_3Asp_4 and CWT are shown at a magnification of 400 000x.

The nature of the fibrous material in the ΔIle_3Asp_4 electron micrograph is unknown.

$\triangle Ile_3 Asp_4$



CWT

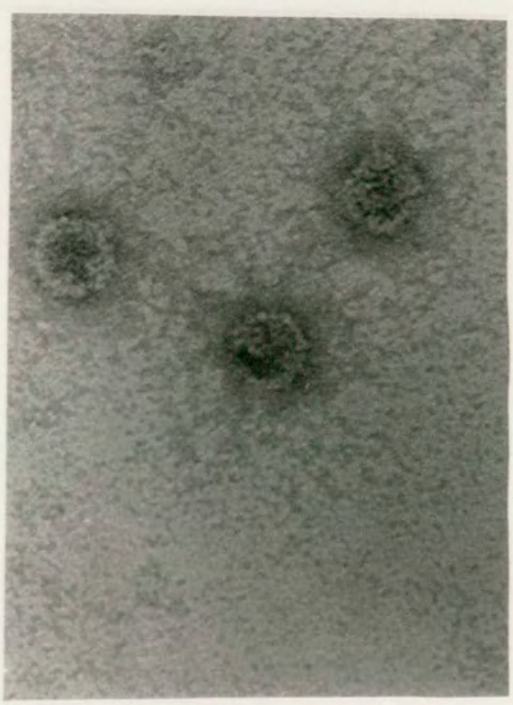
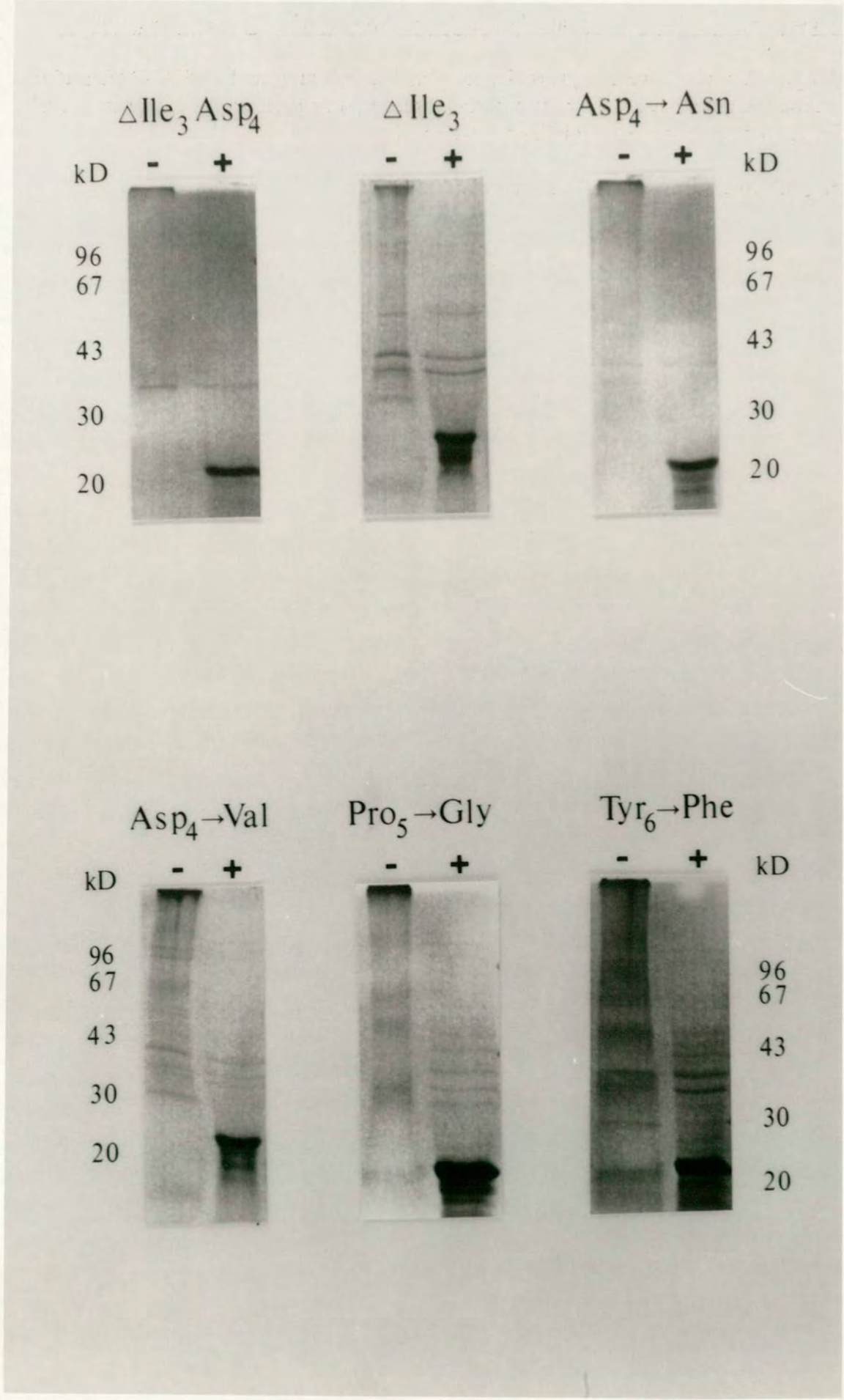


Figure 3.7

Non-reducing SDS-polyacrylamide gel electrophoresis of amino-terminus mutants

All proteins which were expressed were purified and subjected to SDS-PAGE after incubation at 100°C for 5 minutes in the absence (-) or presence (+) of an excess of DTT. Gels were stained with Coomassie Blue.

5μg of protein was loaded in each well.


△Ile3 Asp4
△Ile3
Asp4→Asn
kD
- +
96
67
43
30
20
Asp4→Val
Pro5→Gly
Tyr6→Phe

3.2.7 Radioimmune Assays

The interaction of all purified proteins with antibodies was investigated in a bead-based radioimmune assay. Beads coated with polyclonal anti-HBc/anti-HBe antibodies from human serum were incubated with the following dilutions of purified protein: 5μg of protein was diluted $1/10^3$, $1/5x10^3$, $1/10^4$, $1/5\ x10^4$, $1/10^5$ and $1/10^6$ in BSA solution (RIP Buffer, section 2B.10.3). These beads were then incubated with ^{125}I-labelled polyclonal anti-HBc/anti-HBe IgG and the activity of the bound ^{125}I-IgG measured with a gamma counter. Samples were counted for 10 minutes and a correction factor employed (section 2B.10.5) so that samples measured at different times after the final wash could be directly compared. All samples were assayed in duplicate and the average of the two figures (in counts per minute) for each sample calculated. This value was divided by the average of 6 values obtained from beads which had been processed in an identical manner to the other beads with the exception that no purified protein was added. Therefore, these beads allowed measurement of the background binding activity. This figure was termed the "negative" value, and the number obtained by dividing the sample value (positive, P) by the negative value (N) gave a P/N value which was then plotted on a graph (Figure 3.8).

This graph shows that the activity of pTacpRI-4 (ΔIle_3Asp_4) and that of ΔIle_3 are significantly lower than that of wild-type HBcAg, while the other mutants have activities similar to that of wild-type HBcAg.

3.3 Discussion

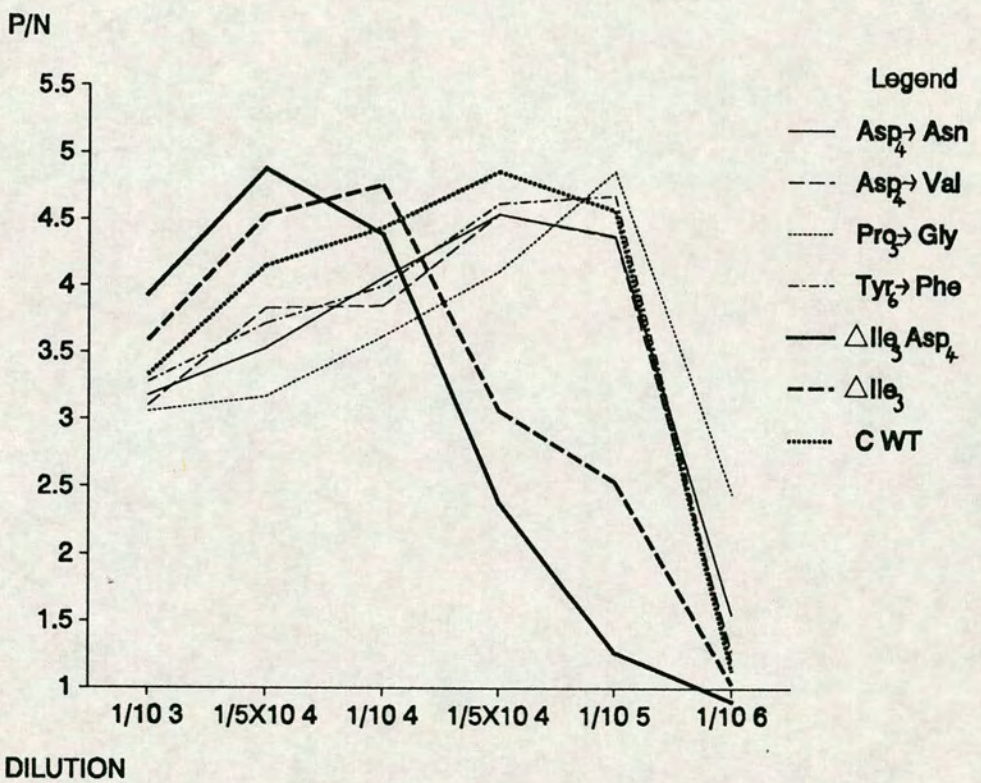
From this work, it is apparent that the reasons for the low activity of pRI-4 (ΔIle_3Asp_4) in the original work by Stahl *et al.* (1982) are two-fold. Firstly, ΔIle_3Asp_4 is present in *E. coli* expressing the appropriate plasmid at a lower level than wild-type HBcAg. Even when the total amount of protein was increased by making use of the *tac* promoter, the relative proportions of wild-type HBcAg and TacpRI-4 (ΔIle_3Asp_4) remained the same and yields of TacpRI-4 (ΔIle_3Asp_4) were consistently less than 10%

Figure 3.8

Radioimmune assay of amino-terminus mutants

5μg of purified protein was diluted by the factors shown on the x-axis of the graph and the radioimmune assay carried out as described in section 2B.10.4, with polyclonal anti-HBc/anti-HBe antibodies.

P/N = positive (sample) cpm / negative control cpm after employment of the correction factor (section 2B.10.5)


P/N
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
1/10 3
1/5X10 4
1/10 4
1/5X10 4
1/10 5
1/10 6
DILUTION
Legend
Asp4→ Asn
Asp4→ Val
Pro3→ Gly
Tyr6→ Phe
△Ile3 Asp4
△Ile3
C WT

of the yield of wild-type HBcAg (from pTacpCore). Similarly, while substitution mutations involving amino acids 5 or 6 ($Pro_5 \rightarrow$ Gly and $Tyr_6 \rightarrow$ Phe) were expressed at levels approximating to wild-type, two substitution mutations involving amino acid 4 were expressed at only approximately 10% of wild-type levels, and expression of the substitution mutation involving amino acid 3 ($Ile_3 \rightarrow$ Asp) could not be detected at all. In contrast, the deletion mutant ΔIle_3 was expressed, but at very low levels (approximately 10% of wild-type), while the expression of ΔAsp_4 and ΔPro_5 could not be detected.

The expression of HBcAg in *E. coli* is apparently highly sensitive to perturbations of sequence. Truncated C genes whose coding sequence terminates at the codon for amino acid 148 are expressed at higher levels than wild-type full-length HBcAg (Figure 4.7), and Nassal *et al.* (1992a), while various deletion mutants of the C gene, involving several different regions of HBcAg, produced no detectable protein (H. Davidson, personal communication), and sequence insertions at the amino terminus of HBcAg often result in a drastic reduction in expression levels (A. Shiau, personal communication), or an inability to express the protein (Clarke *et al.*, 1987) in *E. coli.*

This sensitivity may be either at the RNA level, resulting in reduced production of the mRNA or reduced translation of the mRNA; or at the protein level, whereby the mutant proteins are in fact produced but are more sensitive to protein-degrading elements within the cell and are degraded. In the case of the deletion mutants created by H. Davidson (above), this is thought not to be the case as samples of cells taken immediately after induction of expression with IPTG, and at many time points thereafter, contained no detectable HBcAg (H. Davidson, personal communication). As HBcAg-specific mRNA was detected indirectly in both TG1 and W3110Iq cells containing each mutant plasmid, it is apparent that lack of production of core-specific mRNAor lower stability of this mRNA is not the cause of low detection of the protein products. As the assay employed in this work is not quantitative there may be some variability in the levels of mRNA present, but it is clear that in the cases where no protein can be detected mRNA can be detected. This was also true for the deletion

mutants created by H. Davidson (above) which produced mRNA but no detectable protein (H. Davidson, personal communication).

The arginine-rich carboxy-terminal region of HBcAg produced in *E. coli* is known to bind both DNA and RNA (Matsuda *et al.*, 1988; Gallina *et al.*, 1989; Hatton *et al.*, 1992) and so may not be desirable for the cell. In addition, it may bind HBcAg-specific mRNA, of which there will be large quantities in a cell containing an over-expressing plasmid with HBcAg-encoding sequences. If this arginine-rich region is not present, as in the truncated proteins, no such binding would occur and this may explain the higher level of truncated protein relative to full-length protein found upon expression of the two proteins from the same promoter in the same *E. coli* strains. Alternatively, a limiting factor may be the availability of sufficient transfer RNAs for the large number of arginine residues present at the carboxy-terminal region of the protein. If this is indeed the case, it may be remedied by co-transformation with a plasmid containing the gene (*dnaY*) which encodes arginine tRNAs, thus resulting in the production of an excess of arginine tRNAs. However, the mutations that result in *lower* levels of protein being produced in this work are so varied in terms of both the nature of the mutations and the positions of the mutations within the protein that a similar explanation is less likely.

The results of the radioimmune assay indicate that the low activity of ΔIle_3Asp_4 in the original radioimmune assay (Stahl *et al.*, 1982) may not have been due entirely to a low level of expression of the protein in *E. coli.* The radioimmune assay used by Stahl *et al.* utilised only ammonium sulphate-precipitated extracts. While each sample was used at the same overall concentration, the concentrations of the HBcAg-derived proteins within the extracts may not have been the same. Indeed, this work demonstrates that the amount ofpTacRI-4 product (ΔIle_3Asp_4) present was less than 10% of that of TacpRI-11. The radioimmune assay in this work, in contrast, makes use of purified proteins. As the same amount of each sample was used and each sample was of a very similar level of purity, all the samples can be compared directly.

While the activity of TacpRI-4 (ΔIle_3Asp_4) is not as low, relative to the protein consisting of amino acids 3-183 of HBcAg, in the radioimmune assay in this work as in the original radioimmune assay, for reasons probably relating to expression levels described above, it is still significantly lower than that of the latter protein. Similarly, the mutant in which only Ile_3 is deleted shows activity significantly lower than wild-type HBcAg. In contrast, all substitution mutations examined interacted with antibody to a similar extent to wild-type HBcAg. As deletion of Ile_3 has a similar effect to deletion of Ile_3 and Asp_4 together, and substitution of Asp_4 with either Asn or Val has no effect, it may be that Ile_3 is the more important amino acid in this context. As ΔIle_3Asp_4 has a slightly lower activity than ΔIle_3 it may be that deletion of Asp_4 also contributes to the decrease in antibody interaction. However, this radioimmune assay may not be sufficiently accurate for such a conclusion to be made. It would therefore be useful to achieve expression of ΔAsp_4 and include it in a radioimmune assay. In addition, a radioimmune assay involving monoclonal antibodies may be informative as this may further elucidate the changes in structure caused by mutations at the amino terminus of the protein.

While in the context of the proteins used in this work amino acids 1 and 2 are not important for formation of normal core particles, it would be useful to create constructs in which the first amino acid of the expressed protein is amino acid 1 of HBcAg, with no fusion sequences present. This would determine whether the effect of the deletions described above is peculiar to this protein context or whether it is also true of the protein in its normal context and can therefore be extrapolated to the *in vivo* situation.

CHAPTER 4: Disulphide Bond Formation in Core Particles

4.1 Introduction

HBcAg contains 4 cysteine residues which are completely conserved among mammalian hepadnaviruses (Figure 1.6) and occur at amino acid positions 48, 61, 107 and 183. In contrast, the nucleotide codons for these residues show no preference for T or C in the third position (Figure 4.1). Therefore the conservation appears to be of amino acids rather than of nucleotides and suggests an important function for these cysteines, the most obvious of which would be their participation in disulphide bonds. Disulphide bonds can form between two cysteines only when the amino acids are in the correct orientation and the α-carbon atoms are 0.4-0.9 ηm apart. The presence of an electron acceptor is also essential and so the redox potential of the environment is also important. Disulphide bonds can form between cysteines of the same (intra-molecular) or separate (inter-molecular) polypeptide chains, if all the above criteria are met. As 180 subunits of HBcAg interact to form the nucleocapsid of HBV, it was possible that the cysteine residues of HBcAg could be involved in both intra-molecular (within the same molecule) and inter-molecular (betwen separate molecules) disulphide bonds. It has been noted that, in the absence of reducing agent, core particles produced in *E. coli* will not enter a denaturing polyacrylamide gel, while in the presence of an excess of reducing agent the protein will run in the gel as a 21kD monomer (our laboratory, unpublished observations; Gallina *et al.*, 1989; Jeng *et al.*, 1991). This suggests that intermolecular disulphide bonds are indeed formed.

A straightforward means of determining the nature and importance of these bonds was to mutate the coding sequence of the C gene such that cysteine codons were replaced by codons for another amino acid. The effects of these mutations on the disulphide bond complement of the core particle were then assessed by investigating the mobilities of the mutant proteins in a SDS-polyacrylamide gel as disulphide bonds are covalent and therefore are not broken by the action of denaturing agents. Cysteine codons were replaced by serine codons as the corresponding amino acids are similar in size but serine has an oxygen atom in place of the sulphur atom of cysteine. This precludes serine residues from any involvement in disulphide bonds, but does not

Figure 4.1

Cysteine codons of HBV Core gene

	CYS48	**CYS61**	**CYS107**	**CYS183**
HBV*adyw*	TGT	TGC	TGT	TGT
HBV*adr*	TGT	TGT	TGC	TGT
HBV*adr4*	TGT	TGT	TGT	TGT
HBV*adw*	TGC	TGC	TGC	TGT
HBV*adw2*	TGC	TGC	TGC	TGT
HBV*ayw*	TGT	TGC	TGT	TGT
HBV*ayr*	TGT	TGT	TGC	TGT
WHV	TGC	TGC	TGT	TGC
GSHV	TGT	TGT	TGT	TGC

Overall nucleotide sequence **identity** of the 549 nucleotides of the Core gene:

HBV subtypes, WHV and GSHV - 58.3%

HBV subtypes only - 84.5%

make a gross alteration to the amino acid structure, and potentially therefore, to the overall protein structure. Cysteine residues were mutated to serines singly, in pairs and in trios, and in the context of both the full-length HBcAg and a COOH-terminally-truncated (HBeAg-like) protein in order to gauge any effect the arginine-rich carboxy-terminus may have on the folding of the protein, which may be reflected in the disulphide bond pattern of the two proteins as the structural requirements for disulphide bond formation are so strict.

4.2 Results

4.2.1 Plasmid Constructions

Plasmid pTacpCore contains sequences encoding amino acids 1-8 of β-galactosidase fused to 3 linker amino acids (Glu Phe His) which are in turn fused to amino acids 3-183 of HBcAg. These sequences are expressed under control of the *tac* promoter, which is induced by the presence of IPTG. This plasmid was created by digestion of pRI-11 (section 2A.2.2) with *EcoR*I and *BamH*I, to liberate a 1005 bp fragment containing the coding sequences described above, which was then ligated to the larger (4373 bp) *EcoRI/BamHI* digestion product of pTacHpaIIR2 (section 2A.2.2).] The resulting plasmid was used to express large amounts of wild-type HBcAg in E. coli.

Plasmid pRI-11E contains the coding sequence for amino acids 1-8 of β-galactosidase fused to 3 linker amino acids (Glu Phe His) which are in turn fused to amino acids 3-148 of HBcAg, which are expressed under control of the *tac* promoter. This plasmid was created as shown in Figure 4.2 by using the polymerase chain reaction to amplify the above coding sequences from plasmid pTacpCore, which had been digested with *EcoR*I and *BamH*I and the resulting 1005 bp fragment ligated between the *EcoR*I and *BamH*I sites of M13mp18. The oligonucleotide primers used were

(i) 568C: 5′ - ATA ACA ATT TCA CAC AGG AAA CAG C - 3′

Figure 4.2

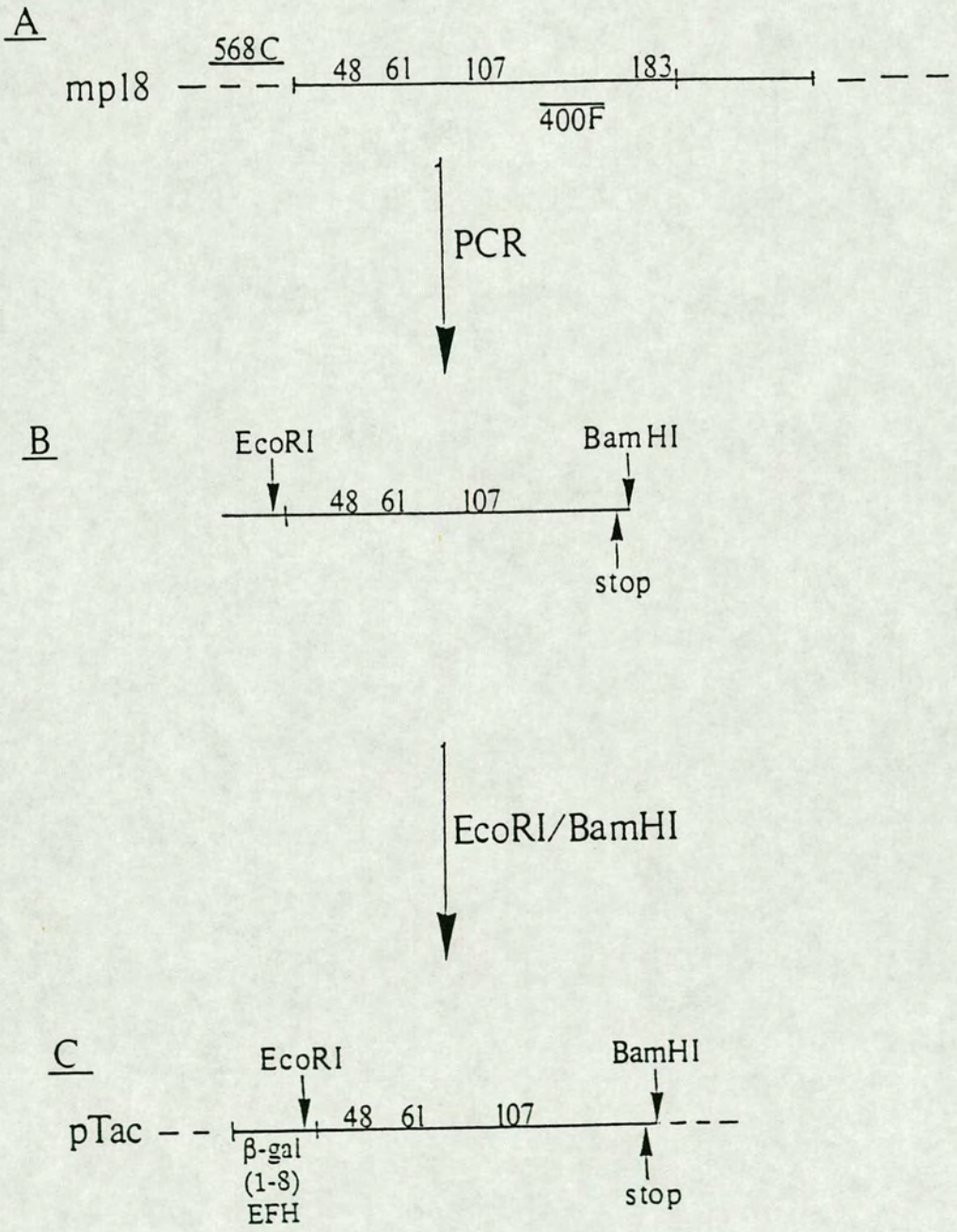
Construction of plasmids encoding truncated proteins

A. DNA encoding the full-length HBcAg was inserted between the *EcoR*I and *BamH*I sites of mp18. The polymerase chain reaction was then performed using primers 568C (section 4.2.1) which hybridises upstream of the *EcoR*I site within mp18; and 400F (section 4.2.1) which hybridises to sequences upstream of the codon for amino acid 144 of HBcAg and includes an in-frame stop codon and a *BamH*I site.

B. The PCR product was then digested with *EcoR*I and *BamH*I

C. and inserted between the *EcoR*I and *BamH*I sites of pTacHpaIIR2 which had been digested with *EcoR*I and *BamH*I, and the 4373bp vector fragment retained. This created a plasmid capable of expression, under control of the *tac* promoter, of amino acids 1-8 of β-galactosidase followed by three linker amino acids (Glu Phe His) followed by amino acids 3 to 148 of HBcAg.

TRUNCATED PROTEINS



which hybridises to sequences in the non-coding strand upstream of the *EcoR*I site of M13mp18; and

(ii) 400F: 5′ - GGG ATC CTC AAA CAG TAG TCT CCG GAA GTG - 3′

which hybridises to sequences within the coding strand up to and including the Val 148 codon, and also contains an in-frame stop codon immediately downstream of this. When both are double-stranded, oligonucleotide 568C contains an *EcoR*I restriction site and oligonucleotide 400F contains a *BamH*I restriction site, so that digestion of the PCR product with *EcoR*I and *BamH*I produced a fragment which was then ligated to the 4373 bp fragment which resulted from *EcoR*I/*BamH*I digestion of pTacHpaIIR2 (section 2A.2.2). The resulting plasmid was used to produce large amounts of wild-type truncated HBcAg in *E. coli.* The sequences of plasmids pTacpCore and pRI-11E were confirmed by DNA sequencing (Figures 4.3 and 4.4, respectively)

4.2.2 Site-Directed Mutagenesis

The procedure employed is shown in Figure 4.5.
The template for mutagenesis was created by insertion of the 1005 bp fragment produced by *EcoR*I/*BamH*I digestion of pTacpCore, between the *EcoR*I and *BamH*I sites of M13mp18 such that single-stranded template contained the coding strand. A Large-Scale Template Preparation (Materials and Methods) of this was made and used in site-directed mutagenesis.

Initially, the following 5 mutants were made and identified by DNA sequencing across the mutated region. The mutants are named C (for full-length Core protein) followed by the number of the codon(s) mutated to Serine (S). The partner of the nucleotide to be mutated is underlined.

Cys48 → Ser (**C48S**)
oligonucleotide : 879D 5′ - GAG GTG AAG AAT GCT CAG G - 3′

Figure 4.3

Nucleotide Sequence of Full-Length Mutants - amino acid codons 48, 61, 107 and 183.

The nucleotide sequence shown is that of the non-coding strand, and reads from bottom to top, 5′ to 3′. All sequences read, from left to right, GATC

In all cases, mutation of cysteine codons to serine codons required substitution of a C in the second nucleotide position with a G, and this is indicated by the presence of *.

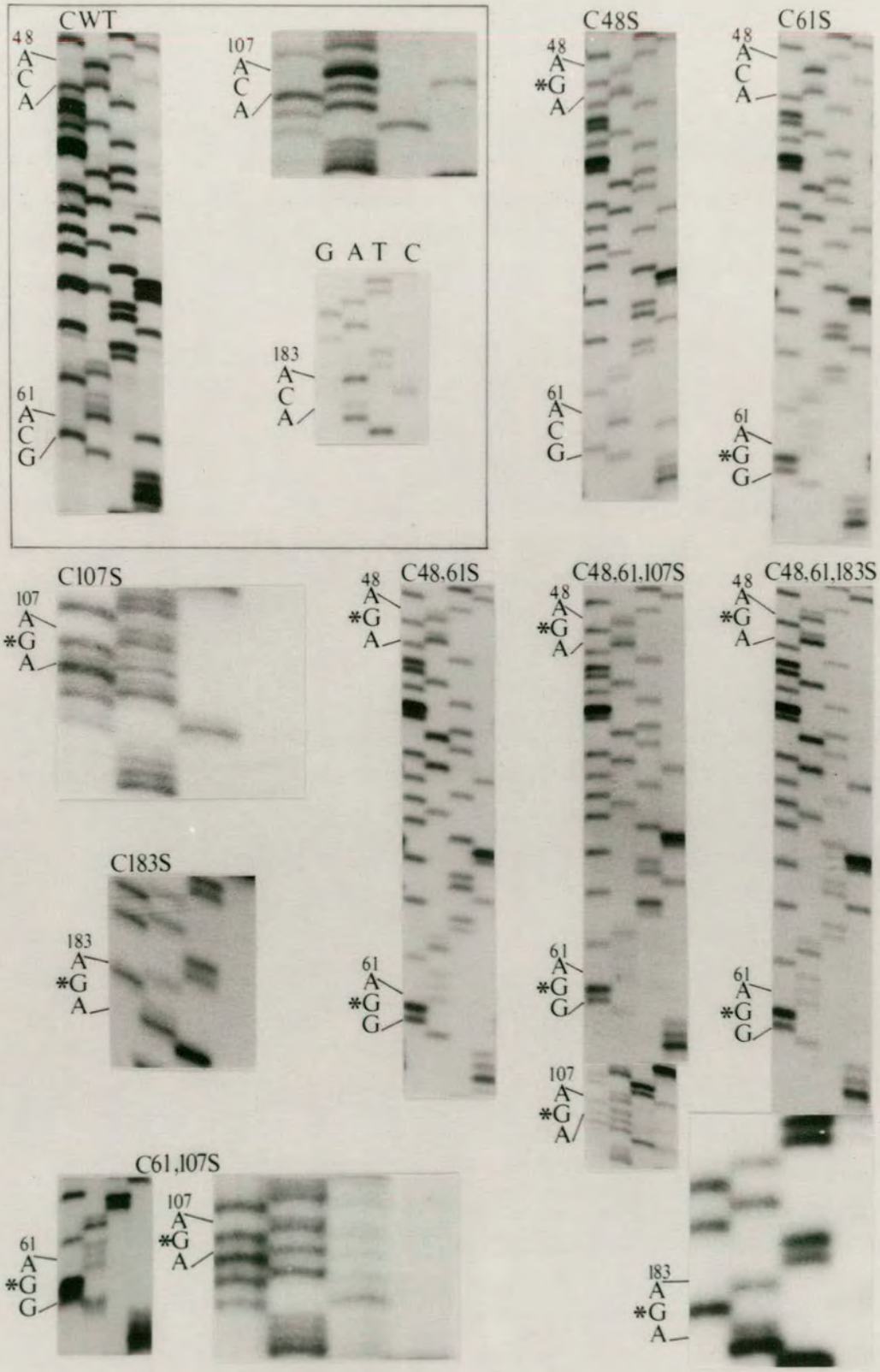

CWT
48
A
C
A
107
A
C
A
G A T C
183
A
C
A
61
A
C
G
C48S
48
A
*G
A
61
A
C
G
C61S
48
A
C
A
61
A
*G
G
C107S
107
A
*G
A
C48,61S
48
A
*G
A
61
A
*G
G
C48,61,107S
48
A
*G
A
61
A
*G
G
107
A
*G
A
C48,61,183S
48
A
*G
A
61
A
*G
G
183
A
*G
A
C183S
183
A
*G
A
C61,107S
61
A
*G
G
107
A
*G
A

Figure 4.4

Nucleotide Sequence of Truncated Mutants - amino acid codons 48, 61 and 107.

The nucleotide sequence shown is that of the non-coding strand, and reads from bottom to top, 5′ to 3′. All sequences read, from left to right, GATC

In all cases, mutation of cysteine codons to serine codons required substitution of a C in the second nucleotide position with a G, and this is indicated by the presence of *.

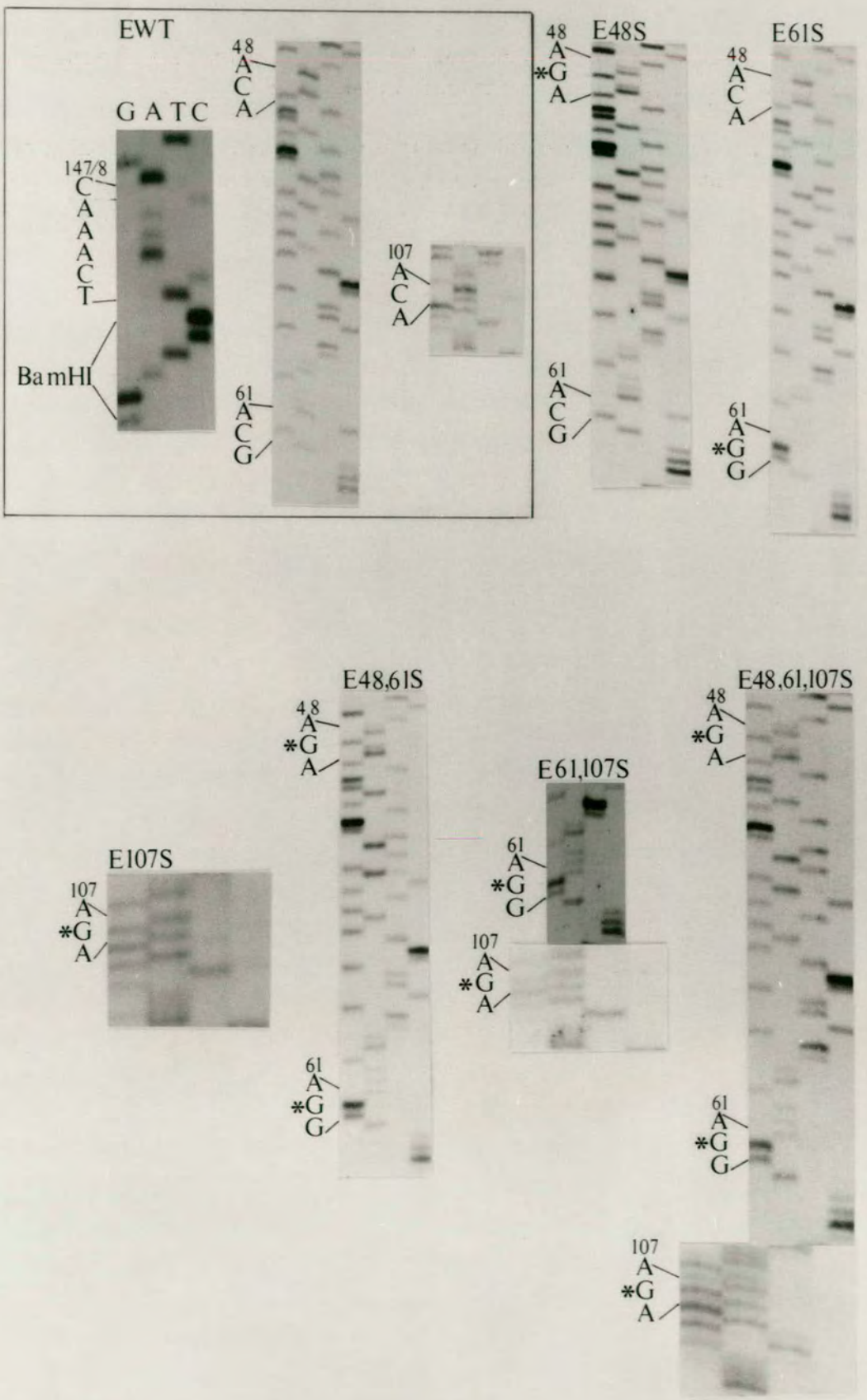

EWT
G A T C
147/8
C
A
A
A
C
T
BamHI
48
A
C
A
61
A
C
G
107
A
C
A
E48S
48
A
*G
A
61
A
C
G
E61S
48
A
C
A
61
A
*G
G
E48,61S
48
A
*G
A
61
A
*G
G
E107S
107
A
*G
A
E61,107S
61
A
*G
G
107
A
*G
A
E48,61,107S
48
A
*G
A
61
A
*G
G
107
A
*G
A

Figure 4.5

Site-directed mutagenesis

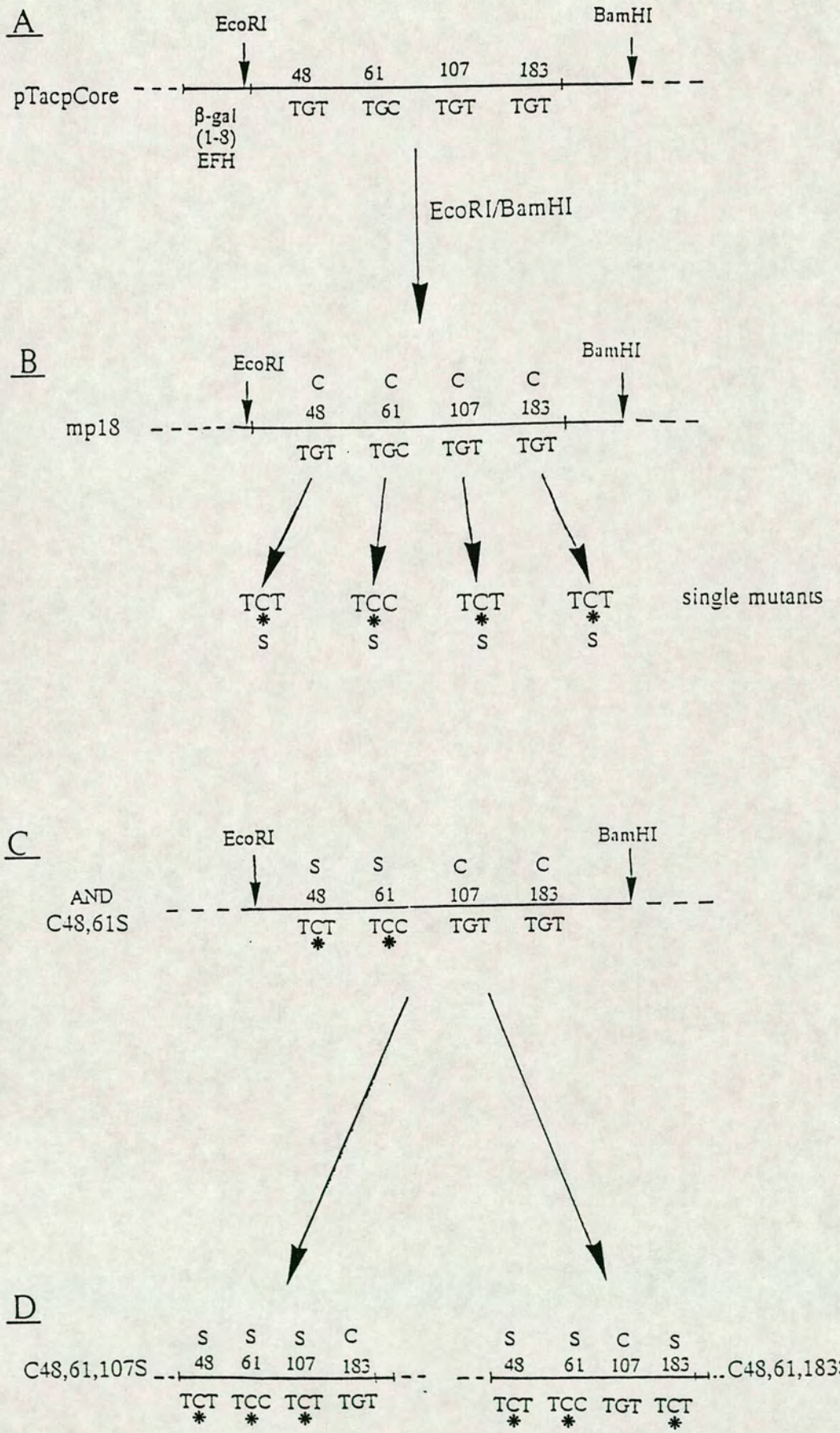
A. The full-length wild-type C gene coding sequence was removed from pTacpCore by digestion with *EcoR*I and *BamH*I.

B. This was inserted between the *EcoR*I and *BamH*I sites of M13mp18 such that single-stranded template contained the coding strand. This template was then subjected to oligonucleotide site-directed mutagenesis, as described in section 2B.8, and initially four single mutants and

C. one double mutant (C48,61S) were created. This double mutant was then used as template for the creation of

D. two triple mutants, C48,61,107S and C48,61,183S, by the same procedure.

The nucleotides which have been changed are indicated by *.


A
EcoRI
BamHI
pTacpCore
β-gal
(1-8)
EFH
48
61
107
183
TGT
TGC
TGT
TGT
EcoRI/BamHI
B
EcoRI
BamHI
mp18
C
C
C
C
48
61
107
183
TGT
TGC
TGT
TGT
TCT
TCC
TCT
TCT
S
S
S
S
single mutants
C
EcoRI
BamHI
AND
C48,61S
S
S
C
C
48
61
107
183
TCT
TCC
TGT
TGT
D
C48,61,107S
S
S
S
C
48
61
107
183
TCT
TCC
TCT
TGT
S
S
C
S
48
61
107
183
TCT
TCC
TGT
TCT
..C48,61,183S

Cys61 → Ser (**C61S**)

oligonucleotide : 880D 5′ - CTC CCC AGG AAA GAA TTG C - 3′

Cys107 → Ser (**C107S**)

oligonucleotide : 881D 5′ - CCA AAA GTG AGA GAA GAA ATG TG - 3′

Cys183 → Ser (**C183S**)

oligonucleotide : 543E 5′ - AGG GAT ACT AAG ATT GAG ATT CC - 3′

Cys48,61 → Ser (**C48,61S**)

oligonucleotides : 879D and 880D (above).

A large scale template preparation of the double mutant C48,61S template was then utilised as template for site-directed mutagenesis to create the mutants:

Cys48,61,107S → Ser (**C48,61,107S**)

Cys48,61,183S → Ser (**C48,61,183S**)

An additional double mutant, Cys61,107 → Ser (**C61,107S**) was created as shown in Figure 4.6, using the polymerase chain reaction. Template C48,61,107S was amplified using primers 568C (see pRI-11E production, above) and the M13 (-20) (New England Biolabs) primer which hybridises to sequences downstream of the *BamH*I site in M13mp18, in the non-coding strand, thereby producing a DNA fragment of 1025bp which contains the entire HBcAg coding region (Product 1). Sequences from template C61S were amplified using primers 568C (above) and

159E: 5′ - GGC CCA CAT TAG TGT TG - 3′

which hybridises to the coding strand approximately 30 nucleotides upstream of the Cys107 codon, thereby producing an amplified DNA product of 325bp (Product 2)

Figure 4.6

Construction of plasmid C61,107S

A. Mutant C61S template in mp18 was amplified using primers 568C (section 4.2.1) which hybridises within mp18 upstream of the *EcoR*I site, and 159E (section 4.2.2) which hybridises upstream of the Cys107 codon.

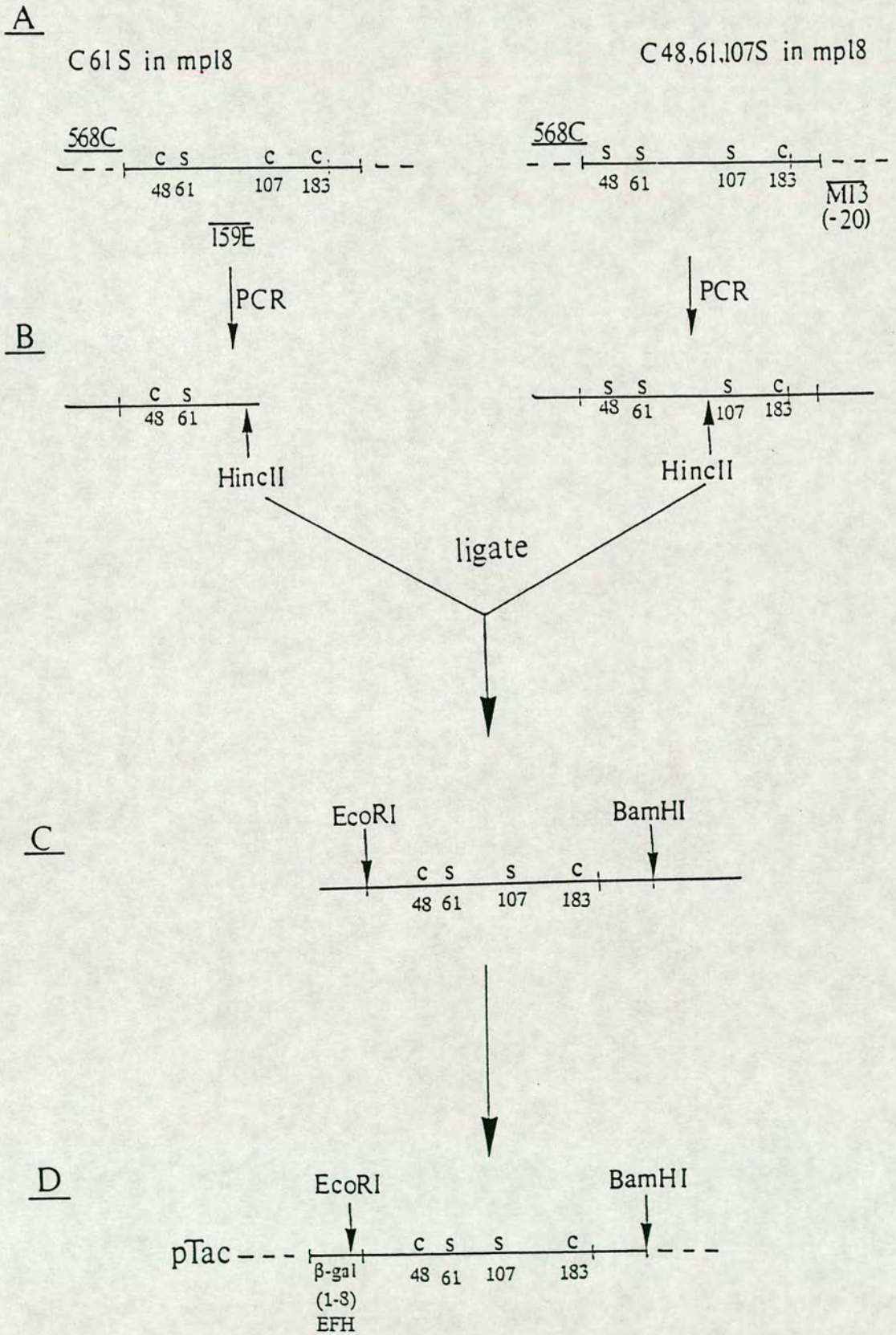
   Mutant C48,61,107S template in mp18 was amplified with 568C (above) and the M13(-20) primer which hybridises to mp18 downstream of the *BamH*I site.

B. PCR products were both digested with *Hinc*II which cleaves the DNA at the 5′ end of the 159E primer region and the digestion product of Product 1, containing Cys48 and Ser61 codons was ligated to the digestion product of Product 2, containing Ser107 and Cys183 codons.

C. This ligated DNA was then digested with *EcoR*I and *BamH*I

D. and inserted into the vector fragment of pTacHpaIIR2 which had been digested with *EcoR*I and *BamH*I. This plasmid has the capacity to encode cysteines at amino acid positions 48 and 183 and serines at positions 61 and 107.

C61,107S construction



which encodes from the amino-terminus up to amino acid 94 of HBcAg. Both PCR products were then digested with *Hinc*II, which recognises a single site within these PCR products and restricts the DNA bluntly one nucleotide from the 5′ end of the 159E primer region. The 269 bp fragment from digestion of Product 2 therefore contains sequences encoding a cysteine at position 48 but a serine at position 61 and was then ligated to the 700 bp fragment of Product 1, which contains sequences encoding a cysteine at position 183 but a serine at position 107. The ligated DNA product therefore encodes cysteines at amino acid positions 48 and 183 but serines at amino acid positions 61 and 107.

All mutants were sequenced along the length of the coding region and are shown in Figure 4.3.

Sequences encoding truncated mutant proteins were produced from the mutants described above by utilising the M13mp18 templates containing the mutated coding sequences as templates for the polymerase chain reaction (Figure 4.2). The procedure and the oligonucleotide primers used were those employed to produce plasmid pRI-11E (above), thereby encoding proteins whose carboxy-terminal amino acid is Val148. Thus the truncated mutant proteins encoded were:

E48S	E48,61S	E48,61,107S
E61S	E61,107S	
E107S		

where "E" indicates "HBe-like" or truncated proteins and the remainder of the nomenclature is the same as that for the full-length proteins.

All PCR products were cloned into M13mp18 via the *EcoR*I and *BamH*I sites, and sequenced. The sequence was correct in all cases and therefore no mistakes had been made by the *Taq* polymerase in copying the coding region and flanking restriction sites (Figure 4.4).

Replicative Form preparations were made of the full length mutants, with the

exception of C61,107S; and these and the PCR product form of the other mutants were digested with *EcoR*I and *BamH*I. The products of this digestion were then ligated to the 4373bp vector fragment released by *EcoR*I/*BamH*I digestion of pTacpCore or pTacHpaIIR2 to form the complete expression plasmids.

4.2.3 Expression and Purification

These plasmids were introduced into *E. coli* strain W3110Iq and expression of the desired proteins induced by addition of IPTG to a final concentration of 50mM. Examination of crude extracts of these cultures showed that all mutants were expressed at a similar level to their wild-type equivalent, but that truncated proteins were expressed at a consistently higher level than the full-length proteins (Figure 4.7). Crude extracts were then subjected to ammonium sulphate precipitation and all HBcAg-derived proteins were precipitated at 35% ammonium sulphate saturation. Passage of dialysed ammonium sulphate precipitated extracts through a Sepharose 4B column resulted in all HBcAg-derived proteins being eluted before fraction 11. This corresponds to a species of very high molecular weight, such as a core particle composed of 180 molecules of full-length protein, which has a predicted molecular weight of (180 x 21kD) = 3.78 x 10^6D; or a core particle composed of 180 molecules of truncated protein, which has a predicted molecular weight of (180 x16kD) = 3.06 x 10^6D. The behaviour of all mutant proteins during purification was therefore consistent with their existence in the form of high molecular weight particles.

4.2.4 Electron Microscopy

When examined with the electron microscope, truncated proteins were shown to exist as particles with morphology indistinguishable from those formed by wild-type full-length proteins. In addition, all mutant proteins, both full-length and truncated, have morphology indistinguishable from wild-type as shown in Figures 4.8 (full-length) and 4.9 (truncated). In addition, all mutants were examined at the same time by D. Notman and were seen to have apparently identical morphology (not shown).

Figure 4.7

Crude and Ammonium Sulphate-Precipitated Extracts of Full-Length and Truncated Proteins

Full-length protein = C61S

Truncated protein = EWT

Samples were incubated at 100°C for 5 minutes in the presence of an excess of DTT and immediately electrophoresed through a 15% SDS-Polyacrylamide gel.

crude = supernatant after lysis and centrifugation of cells (section 2B.9.1)

$(NH_4)_2SO_4$ = sample after precipitation with ammonium sulphate at 35% saturation and dialysis against 50mM Tris pH 8.0 (section 2B.9.2)

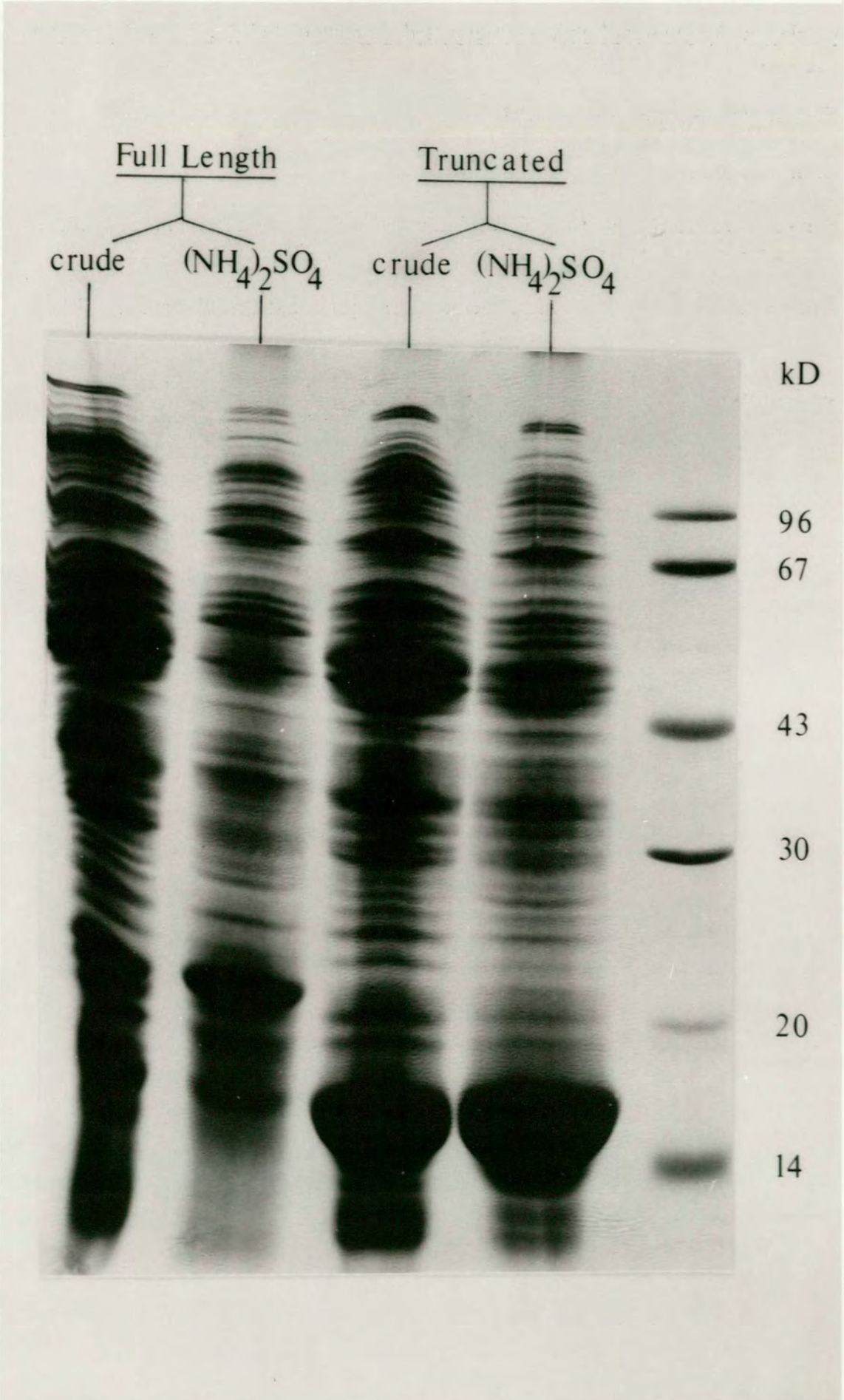

Full Length
Truncated
crude
$(NH_4)_2SO_4$
crude
$(NH_4)_2SO_4$
kD
96
67
43
30
20
14

Figure 4.8

Electron Micrographs of Full-length Proteins

Proteins CWT, C48S, C61S, C107S, C183S and C48,61S:

Magnification = 124 000x

Microscopy by P. Highton

Proteins C48,61,107S and C48,61,183S:

Magnification = 75 000x

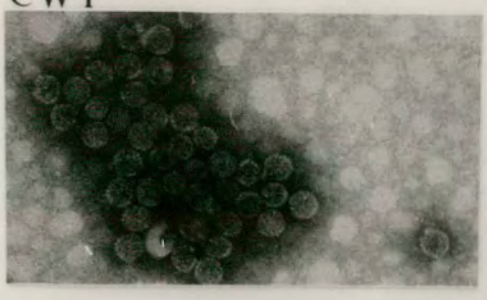
Microscopy by D. Notman

Proteins CWT and C61,107S:

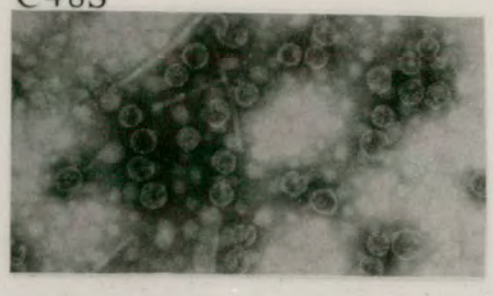
Magnification = 500 000x

Microscopy by S. Bury

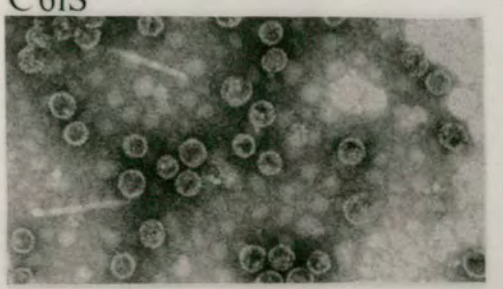
CWT



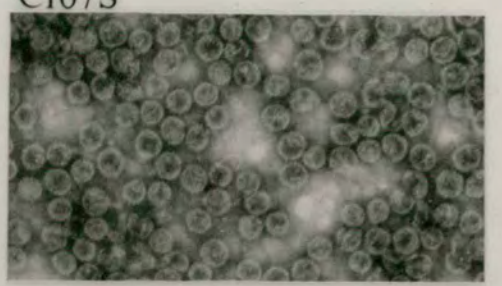
C48S



C61S

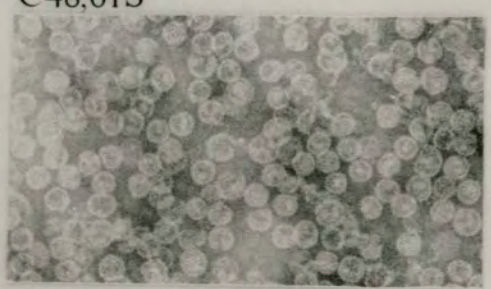


C107S

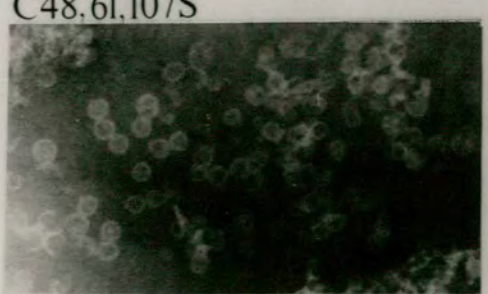


C183S

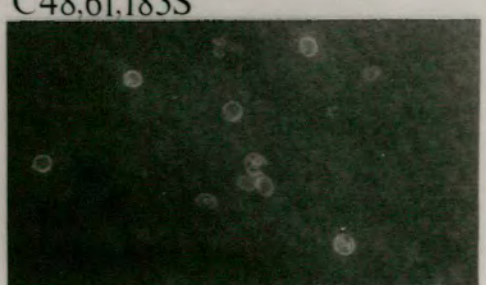
C48,61S



C48,61,107S



C48,61,183S



CWT

C61,107S

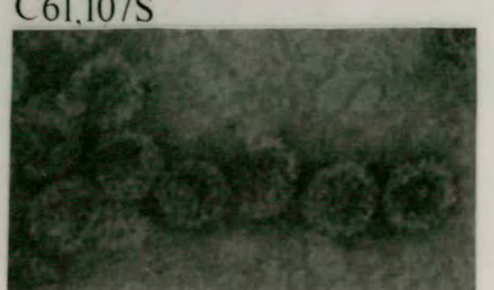


Figure 4.9

Electron Micrographs of Truncated Proteins

Proteins EWT, E48S, E61S, E107S, E48,61S and E48,61,107S:

Magnification = 75 000x

Microscopy by D. Notman

Protein CWT:

Magnification = 315 000x

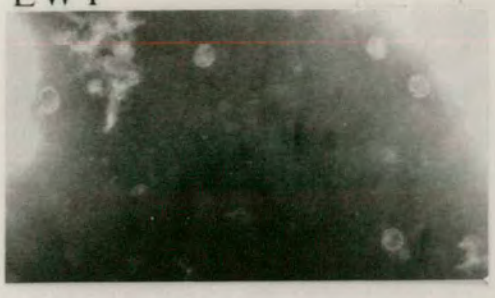
Microscopy by S. Bury

Protein E61,107S:

Magnification = 250 000x

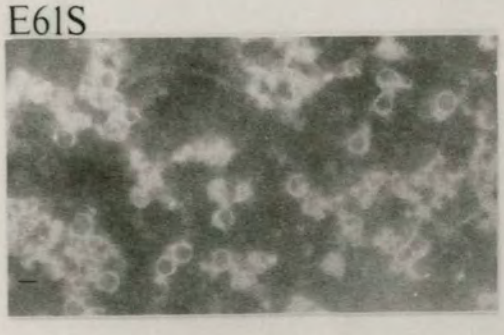
Microscopy by S. Bury

EWT

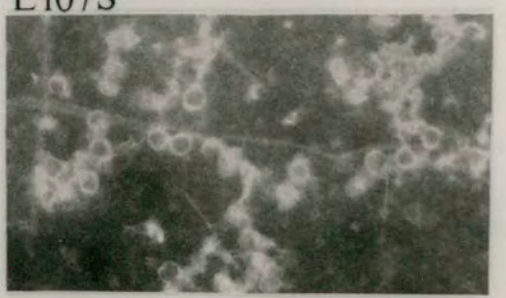


E48S

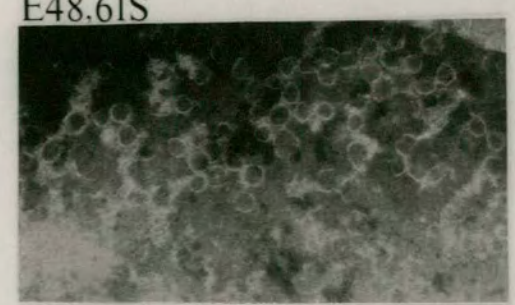
E61S



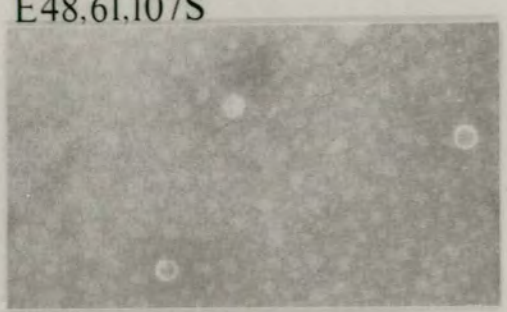
E107S



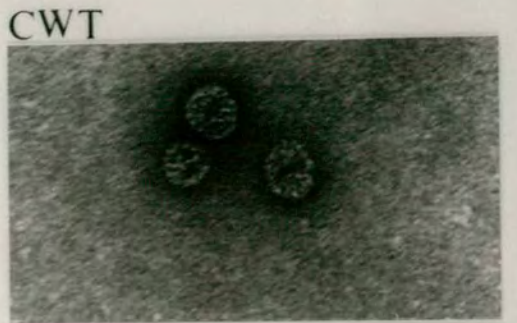
E48,61S



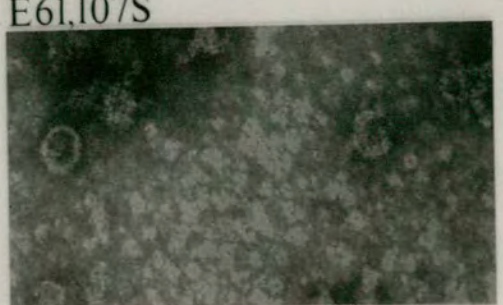
E48,61,107S



CWT



E61,107S



From these photographs it is apparent that both full-length and truncated proteins produced in this expression system can form apparently normal particles and that replacement of cysteines with serines does not alter the structure of the protein in such a way as to prevent full-length or truncated proteins from forming particles of normal morphology. It can therefore be concluded that "normal" core particles can be formed even from full-length HBcAg containing only one cysteine, and from a truncated protein which contains no cysteines, demonstrating that cysteine residues and any disulphide bonds they may form are not essential for core particle formation.

4.2.5 Radioimmune Assays

The interaction of all purified proteins with antibodies was investigated in a bead-based radioimmune assay. Beads coated with polyclonal anti-HBc/anti-HBe antibodies were incubated with the following dilutions of purified protein: 5μg of protein was diluted $1/10^3$, $1/5x10^3$, $1/10^4$, $1/5x10^4$, $1/10^5$, $1/10^6$ in BSA solution (RIP Buffer, section 2B.10.3). These beads were then incubated with ^{125}I-labelled polyclonal anti-HBc/anti-HBe IgG (section 2B.10.2) and the activity of the bound ^{125}I-IgG measured with a gamma counter. Samples were counted for 10 minutes and a correction factor employed (section 2B.10.5) so that samples measured at different times after the final wash could be directly compared. All samples were assayed in duplicate and the average of the two figures (in counts per minute) for each sample calculated. This value was then divided by the average of 6 values obtained from beads which had been processed in an identical manner to the other beads with the exception that no purified protein was added. Therefore these beads allowed measurement of the background binding activity only. This figure was termed the "negative" value, and the number obtained by dividing the sample value (positive, P) by the negative value (N) gave a P/N value which was then plotted on graphs (Figures 4.10A and 4.11A). Three samples, C48,61S; C48,61,107S and E48,61,107S were not included in the initial assay. These samples were assayed at a later date, in conjunction with the same positive and negative controls used in the initial assay, and these results are presented in Figures 4.10B and 4.11B.

Figure 4.10

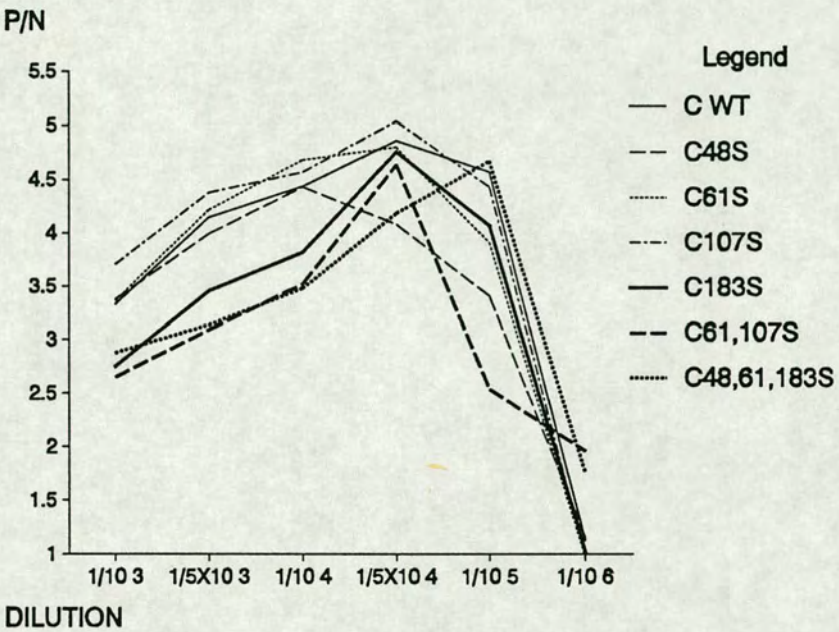
Results of Radioimmune Assays of Full-Length Proteins

P/N = sample activity (positive, P) divided by negative control (N) activity (dilution buffer only), after the correction factor for decay during counting of samples had been applied (section 2B.10.5)

Dilution Values refer to the dilution of 5μg of purified protein by the value shown on the x-axis of the graphs (i.e. $1/10^3$, $1/5x10^3$, $1/10^4$, $1/5x10^4$, $1/10^5$ and $1/10^6$)

All points on the graph are the average of duplicate samples and all samples shown on the same graph were assayed at the same time.

A


P/N
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
1/10 3
1/5X10 3
1/10 4
1/5X10 4
1/10 5
1/10 6
DILUTION
Legend
C WT
C48S
C61S
C107S
C183S
C61,107S
C48,61,183S

B

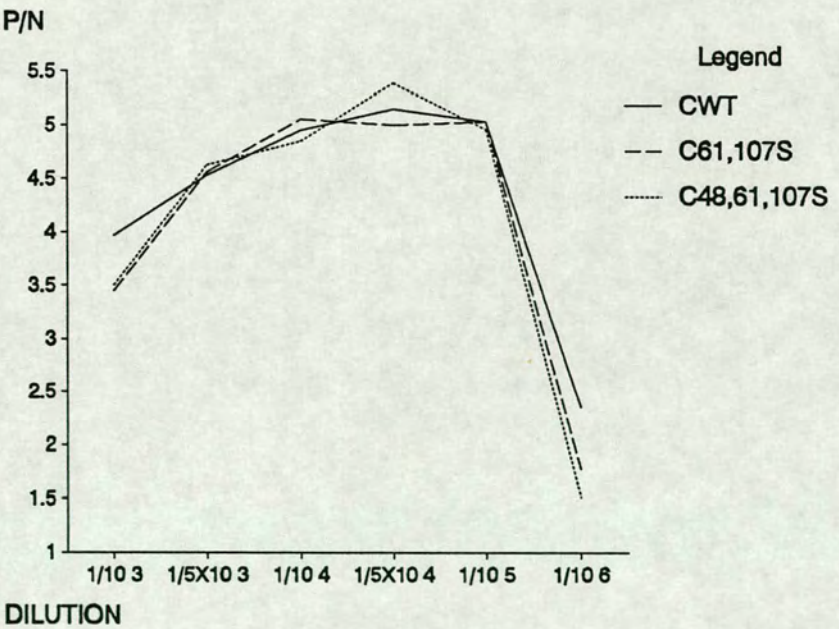

P/N
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
1/10 3
1/5X10 3
1/10 4
1/5X10 4
1/10 5
1/10 6
DILUTION
Legend
CWT
C61,107S
C48,61,107S

Figure 4.11

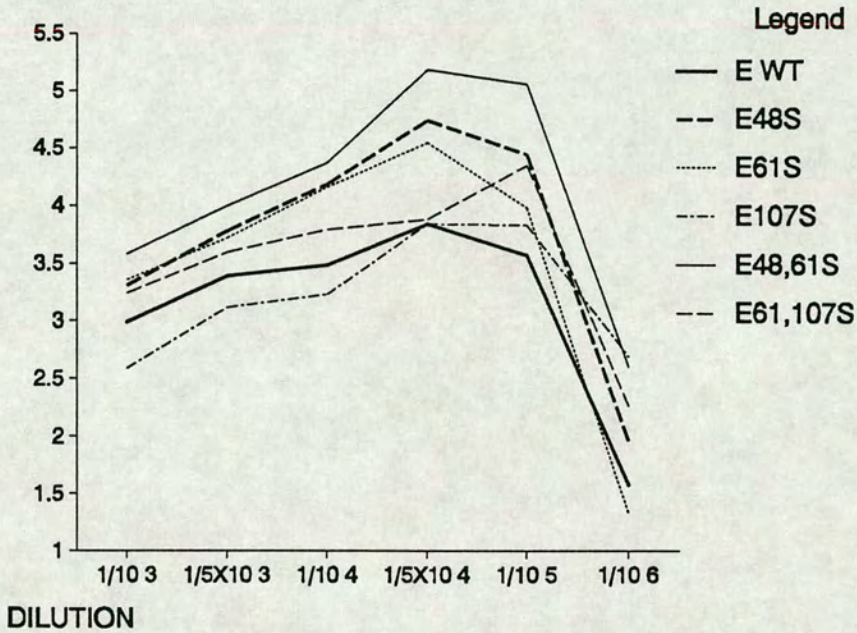
Results of Radioimmune Assays of Truncated Proteins

P/N = sample activity (positive, P) divided by negative control (N) activity (dilution buffer only), after the correction factor for decay during counting of samples had been applied (section 2B.10.5)

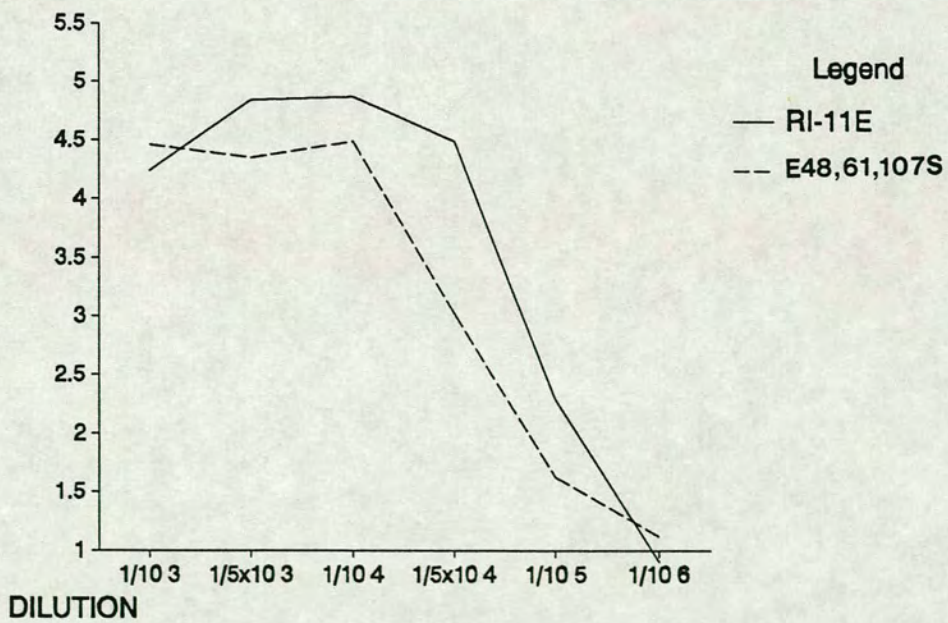
Dilution Values refer to the dilution of 5μg of purified protein by the value shown on the x-axis of the graphs (i.e. $1/10^3$, $1/5x10^3$, $1/10^4$, $1/5x10^4$, $1/10^5$ and $1/10^6$)

All points on the graph are the average of duplicate samples and all samples shown on the same graph were assayed at the same time.

A


P/N
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
1/10 3
1/5X10 3
1/10 4
1/5X10 4
1/10 5
1/10 6
DILUTION
Legend
E WT
E48S
E61S
E107S
E48,61S
E61,107S

B


P/N
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
1/10 3
1/5x10 3
1/10 4
1/5x10 4
1/10 5
1/10 6
DILUTION
Legend
RI-11E
E48,61,107S

These graphs show that the reaction of all mutant proteins with the polyclonal anti-HBc/anti-HBe antibodies is broadly similar to the reaction of these antibodies with wild-type full-length or wild-type truncated HBcAg, and that the reactivity of the truncated proteins and full-length proteins is comparable. Thus the presence of serines in place of cysteines, and the varying numbers of disulphide bonds within these particles (see below) has not altered their structure in such a way that a large difference in affinity for anti-HBc/anti-HBe antibodies has resulted.

4.2.6 Non-Reducing Gel Electrophoresis of Full-Length Proteins

Each full-length protein was electrophoresed through a 15% polyacrylamide gel after incubation in either the presence or absence of an excess of reducing agent (DTT), at the following temperatures for 10 minutes: 0, 20, 40, 60, 80 and 100°C. As shown in Figure 4.12, the gels were dried between two sheets of cellulose membrane backing (Biorad) and scanned with a densitometer. In order to compensate for any differences in loading or staining between tracks or between gels, the densities of bands measured by the densitometer are expressed as % monomer and % dimer. In the absence of DTT, the same results were achieved at all temperatures and representative Coomassie-stained gels are shown in Figure 4.13A. The results of densitometric scanning are shown in Figure 4.13B.

As the triple mutant C48,61,183S (which contains only Cys107) runs as 100% monomer, it can be concluded that Cys107 does not form a bond with a Cys 107 on another molecule. However, as C48,61,107S (which contains only Cys183) exists as 100% disulphide-linked dimer, it can be concluded that a Cys183-Cys183 bond is present

Figure 4.12

Densitometry of Coomassie-stained protein gels

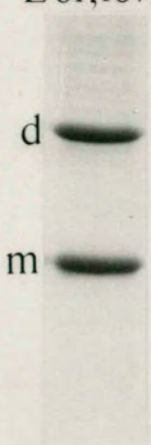
A. A Coomassie-stained, non-reducing, denaturing gel (in this case of E61,107S after incubation with 50μM Cu^{2+}) was dried between two sheets of cellophane membrane backing (Biorad).

B. Scanning by a densitometer of the absorbance at 550ηm along the gel track in A (above) resulted in the production of a graph with two peaks, each corresponding to a band on the gel, which in turn corresponds to either the monomer (m) or dimer (d) form of the protein. The units on the y axis are arbitrary.

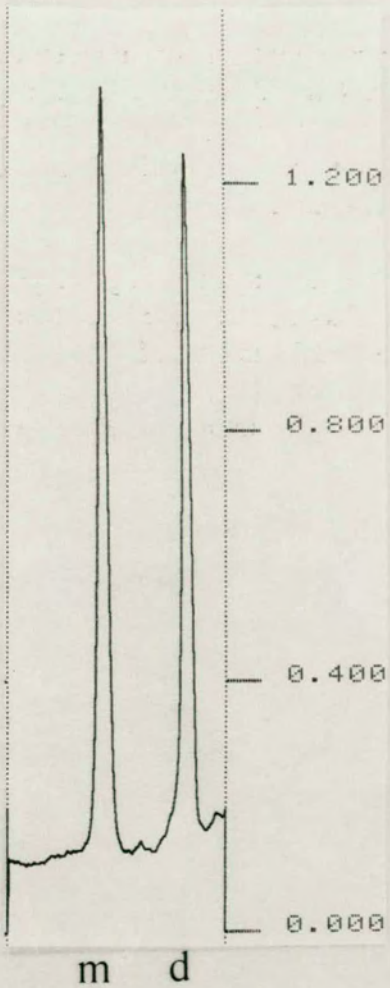
C. The area under each graph peak was measured by the densitometer, in arbitrary units (u), and these were then expressed as a percentage of the total area under all peaks measured.

Thus the amount of each protein species can be expressed as a percentage of the total protein present in each lane of the gel.

A



B



C

Peak	Area	%
m	9832.69u	52%
d	9047.80u	48%

Figure 4.13

SDS-Polyacrylamide Gel Electrophoresis of Full-Length Proteins Under Non-Reducing Conditions

A.

Approximately 5μg of each sample was incubated at 20°C for 10 minutes followed by addition of an equal volume of Sample Buffer (section 2B.9.3) and electrophoresis through a 15% SDS-Polyacrylamide gel.

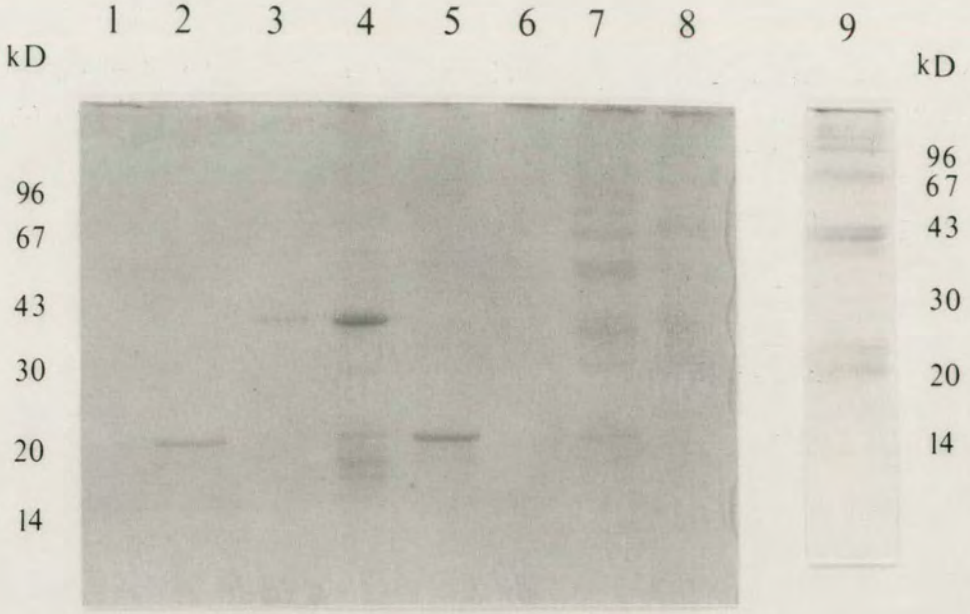
Samples:

1. CWT
2. C48,61,183S
3. C48,61,107S
4. C48,61S
5. C183S
6. C107S
7. C61S
8. C48S
9. C61,107S

B.

Gels were dried between two sheets of cellophane membrane and analysed with a densitometer. The absorbance at 550ηm was measured along each gel track and the areas under the peaks of the resulting graphs measured by the densitometer. These figures were then converted to % monomer and % dimer (Figure 4.12)

A

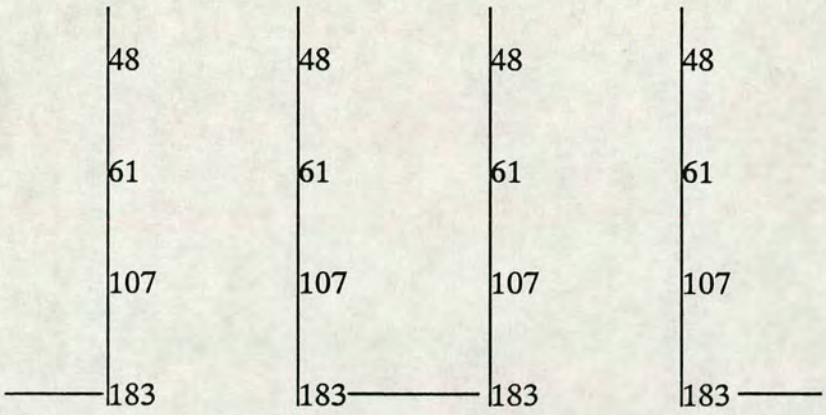


B

Densitometry

1.	CWT	does not enter gel
2.	C48,61,183S	100% monomer
3.	C48,61,107S	100% dimer
4.	C48,61S	86% dimer, 14% monomer
5.	C183S	100% monomer
6.	C107S	does not enter gel
7.	C61S	multimers
8.	C48S	does not enter gel
9.	C61,107S	multimers

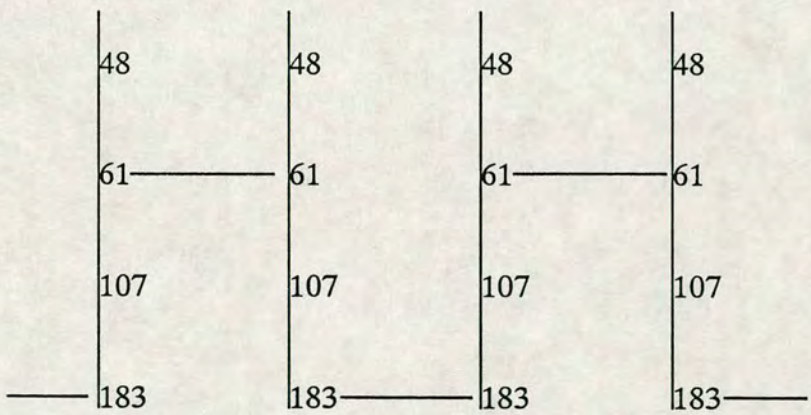
i.e.



These conclusions are confirmed by the observation that C48,61S (which contains both Cys107 and Cys183) also exists as 100% dimer.

A comparison of C48S and C107S shows that neither enter the gel and that they have two cysteines in common - Cys61 and Cys183. It is therefore possible that a bond exists between pairs of cysteine 61s on two monomers, but not between the pairs which are involved in a Cys183-Cys183 bond, thereby creating a disulphide-linked network which cannot enter the gel.

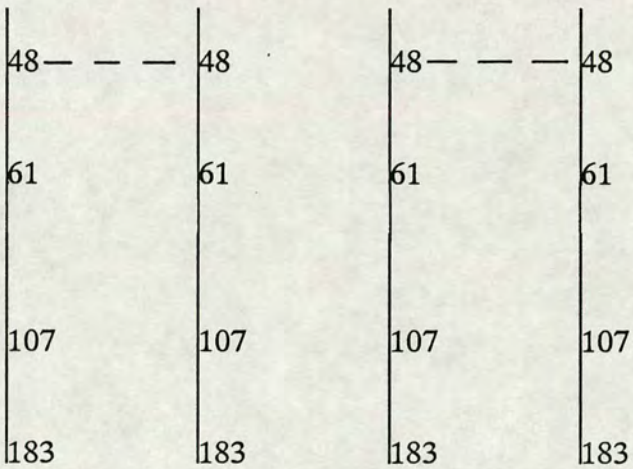
i.e.



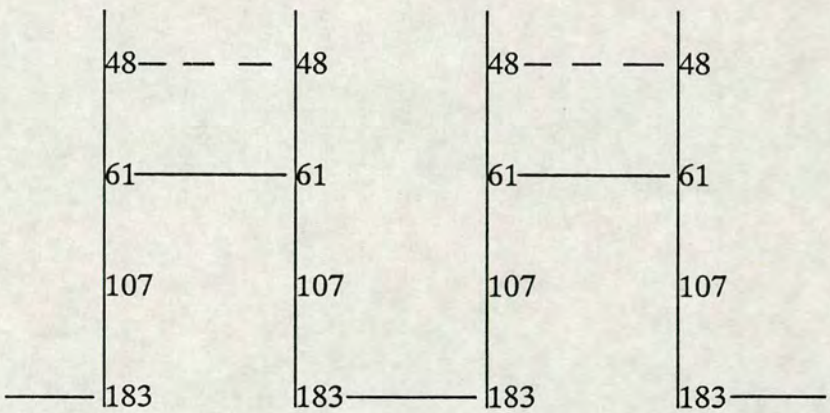
The participation of Cys48 in disulphide bond formation can be deduced by examination of C61S and C61,107S, both of which run in the gel as a ladder of multimers. This suggests that the dimers formed by Cys183 are linked by a Cys48-Cys48 bond (as Cys48 is the only other cysteine present in C61,107S, and Cys107 is thought not to participate in disulphide bonds) but that this bond is not present in

100% of cases. The Cys48-Cys48 bond is also necessarily present between the same pairs of monomers as Cys61-Cys61 and not Cys183-Cys183.

i.e.



Thus the model at this stage is



Several points should be made here. Firstly, while the above model is not the only one that would explain the results obtained with individual proteins, it is the only simple model that can accomodate all the results when taken together (except C183S - see below). Secondly, this assay investigates only the presence of inter-molecular disulphide bonds and so the presence of intra-molecular disulphide bonds cannot be predicted. This question will be addressed in section 4.2.9. Thirdly, the behaviour of the mutant C183S is curious as despite containing cysteines 48, 61 and 107, it exists only as monomers and not the dimers which the above model predicts. As there is no model that can accomodate this result along with the results obtained for all other

mutants, it was possible that disulphide bond formation was not complete at this point and this was investigated as described below.

4.2.7 Oxidation of Full-Length Proteins

1. Air Oxidation

Approximately 9 months after the samples were prepared, during which time they had been stored at -20°C, the gel assay described above was repeated. The mutant C183S, which had previously run in the gel as 100% monomer (Figure 4.14A), now ran in the gel as 93% dimer and 7% monomer (Figure 4.14C). The behaviour of all other proteins was unchanged (results not shown).

2. Cu^{2+} ions

The formation of disulphide bonds requires not only that the two cysteines involved must be positioned correctly, but also that an electron acceptor must be present. The role of electron acceptor is usually taken by O_2 but the reaction is slow and unreliable and is not well characterised. It is apparently dependent upon metal ions, such as Cu^{2+}, which bind transiently to the thiol (SH) group. As these metal ions act catalytically at low concentrations it was possible that incubation of the proteins in the presence of low concentrations of Cu^{2+} would promote disulphide bond formation. Samples were therefore incubated in the presence of 50μM $Cu^{2+}(SO_4)^{2-}$ at room temperature for 20 minutes, before addition of an excess of EDTA and electrophoresis through a 15% SDS-polyacrylamide gel as before.

The original C183S sample, now 93% disulphide-linked dimer after 9 months air oxidation (Figure 4.14C), became 100% dimer after incubation with 50μM Cu^{2+} (Figure 4.14D). Similarly, a fresh sample of C183S was 100% monomer in the absence of Cu^{2+} (Figure 4.14A), but 100% dimer after incubation with 50μM Cu^{2+} (Figure 4.14B). All other proteins showed no altered behaviour after incubation with

Figure 4.14

C183S

All samples were electrophoresed through 15% SDS-polyacrylamide gels under non-reducing conditions. The gels were then dried between two sheets of cellophane membrane and analysed with a densitometer. The absorbance at 550η was measured along each gel track and the areas under the peaks of the resulting graphs calculated by the densitometer. These were then converted to % monomer and % dimer and are shown alongside the appropriate band in the gel.

m = monomer

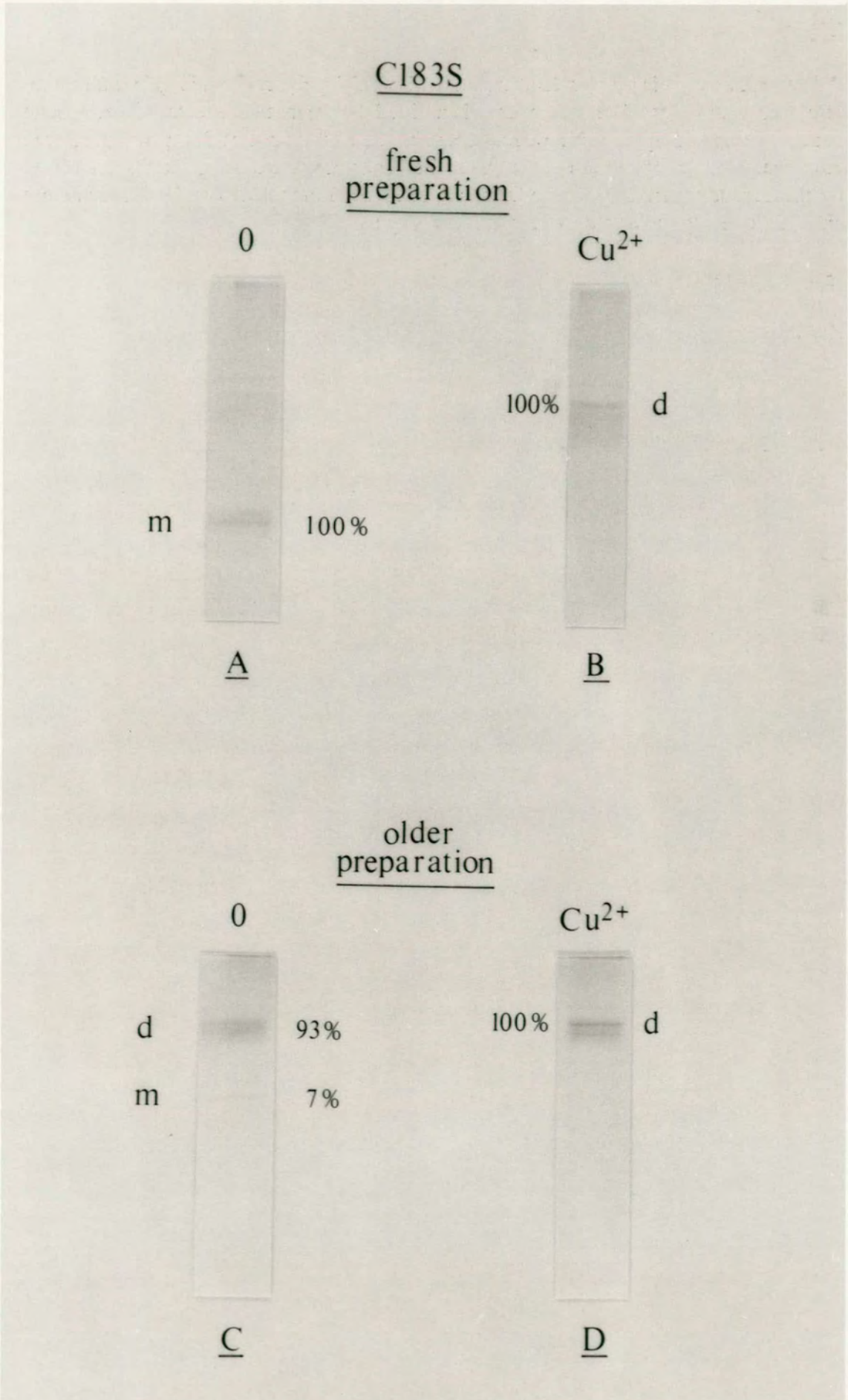
d = dimer

A. Fresh sample

B. Fresh sample immediately following 20 minute incubation with 50μM Cu^{2+}

C. Sample 9 months after preparation

D. Sample 9 months after preparation and immediately following 20 minute incubation with 50μM Cu^{2+}.


C183S
fresh preparation
0
Cu2+
100%
d
m
100%
A
B
older preparation
0
Cu2+
d
93%
100%
d
m
7%
C
D

50μM Cu^{2+} (results not shown).

Thus a more oxidising environment allowed further disulphide bond formation to occur in C183S, such that either or both of the Cys48-Cys48 and Cys61-Cys61 disulphide bonds predicted by the above model could have been formed.

4.2.8 Non-Reducing Polyacrylamide Gel Electrophoresis of Truncated Proteins

The procedure followed was that for the full-length proteins. In addition, proteins were tested approximately 6 months after preparation, and also after incubation in the presence of 50μM Cu^{2+} (both fresh and 6 month samples). The results of this are presented in Figure 4.15 and summarised in Figure 4.16.

The double mutant E48,61S which contains only cysteine 107 exists only as monomer and so a Cys107-Cys107 bond is not present. As in the full-length protein, cysteine 61 is capable of forming a Cys61-Cys61 bond and does so between up to 100% of monomers (E48S). In contrast, the double mutant E61,107S, which contains only cysteine 48 exists as approximately 50% monomer and 50% disulphide-linked dimer. This is attributed to a Cys48-Cys48 bond which, even after exposure to oxidising conditions, forms between only 50% of monomers. As the proteins in which both Cys 48 and Cys 61 are present (EWT, E107S) exist as disulphide-linked dimers and not higher multimer forms, it is apparent that the Cys48-Cys48 and Cys61-Cys61 bonds are between the same monomer. So the model for the truncated proteins is

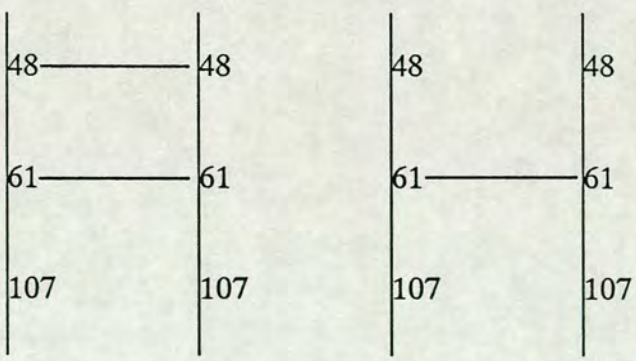


Figure 4.15

SDS-Polyacrylamide Gel Electrophoresis of Truncated Proteins Under Non-Reducing Conditions

A.

Approximately 5μg of each sample (except in lane 2 where 1μg was used) was incubated at 100°C for 10 minutes followed by addition of an equal volume of Sample Buffer (section 2B.9.3) and electrophoresis through a 15% SDS-Polyacrylamide gel.

Samples:
1. EWT
2. E48,61,107S
3. E48,61S
4. E107S
5. E61S
6. E48S
7. E61,107S

B.

Samples were electrophoresed as described for A.

a. samples 6 months after preparation.

b. samples 6 months after preparation, immediately following 20 minute incubation with 50μM Cu^{2+}.

c. freshly-prepared samples following 20 minute incubation with 50μM Cu^{2+}.

m = monomer

d = dimer

N.B. E61,107S. The behaviour of only freshly-prepared samples of this mutant were investigated as the mutant was created more recently and so only fresh preparations were available.

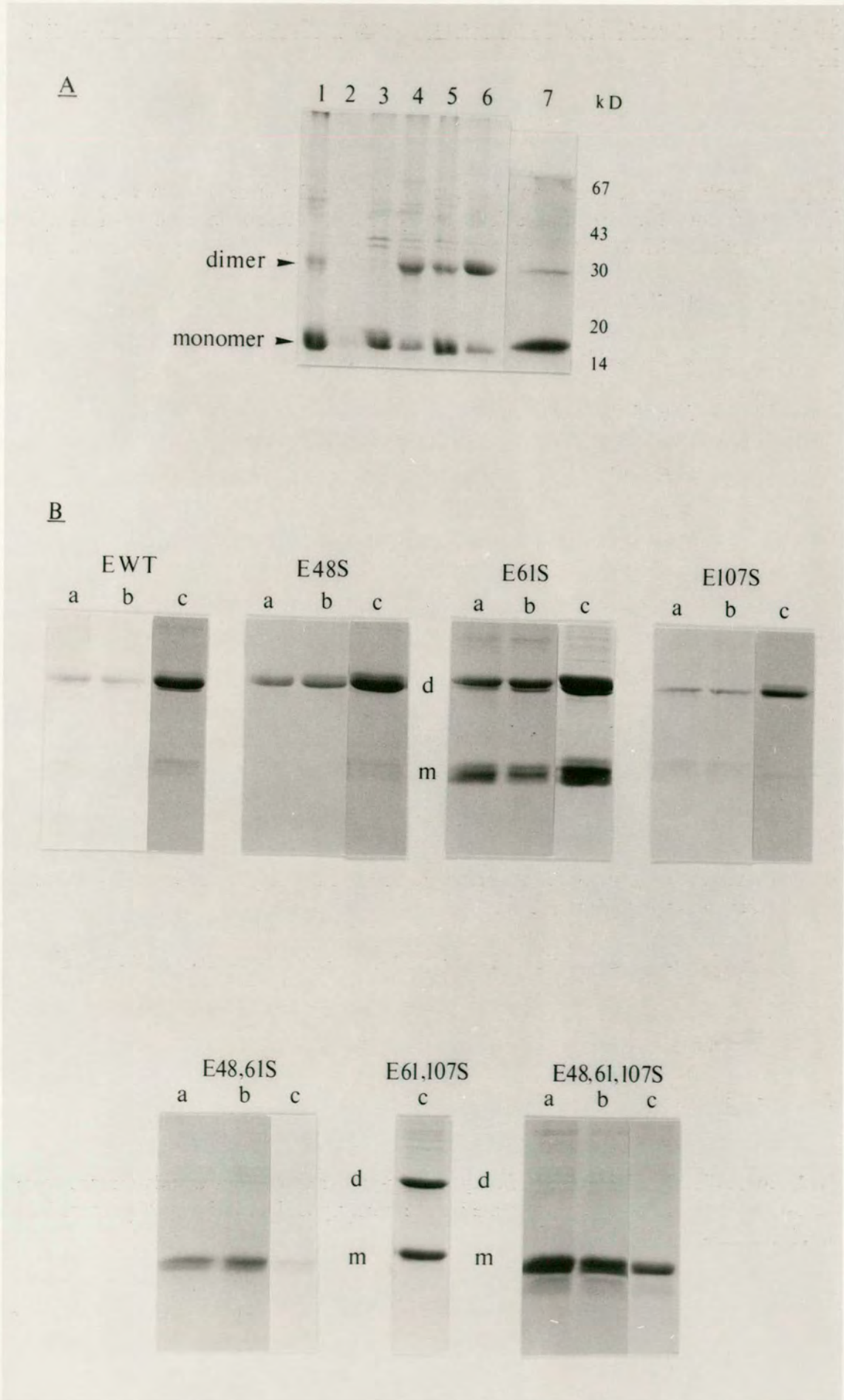

A
1 2 3 4 5 6 7 kD
67
43
30
20
14
dimer
monomer
B
EWT
E48S
E61S
E107S
a b c
d
m
E48,61S
E61,107S
E48,61,107S

Figure 4.16

Densitometric Analysis of Gels (Figure 4.15) of Truncated Proteins

The gels were dried between two sheets of cellophane membrane and analysed with a densitometer. The absorbance of each gel track at 550ηm was measured and the areas under the peaks of the resulting graphs calculated. These were then converted to % monomer and % dimer.

Protein	Cysteines Present	Original	Air	Fresh + Cu^{2+}	Air + Cu^{2+}
EWT	48,61,107	15% d	79% d	93% d	100% d
E48S	61,107	59% d	100% d	96% d	100% d
E61S	48,107	25% d	41% d	48% d	57% d
E107S	48,61	55% d	77% d	100% d	100% d
E48,61S	107	100% m	100% m	100% m	100% m
E61,107S	48	11% d	ND	48% d	ND
E48,61,107S		100% m	100% m	100% m	100% m

Original = fresh preparations, no Cu^{2+} (Figure 4.15A)

Air = following extended air oxidation (Figure 4.15B)

Fresh + Cu^{2+} = fresh preparations incubated with 50μM Cu^{2+} (Figure 4.15Bc)

Air + Cu^{2+} = air oxidised preparations incubated with 50μM Cu^{2+} (Figure 4.15Bb)

ND = no data, as E61,107S prepared only recently

m = monomer

d = dimer

the same as that deduced for these three cysteines in the context of the full-length proteins. However, while the behaviour of the full-length proteins indicated only that the Cys48-Cys48 bond was formed between less than 100% of molecules, the behaviour of the truncated E61,107S and E61S proteins demonstrate that the bond is formed between approximately 50% of molecules.

4.2.9 Is there an intra-molecular disulphide bond in HBcAg?

During investigation of the effect of Cu^{2+} ions on disulphide bond formation, several concentrations of Cu^{2+} were employed. Those proteins that did not enter the gel even in the absence of Cu^{2+} were not examined as further disulphide bond formation would not be detected by this gel electrophoresis assay. At the lowest concentration utilised (50μM) the results were as described above and were consistent with the results of prolonged air oxidation of the proteins. However, at higher Cu^{2+} concentrations (the highest utilised was 50mM) larger multimers were formed. This was true both of fresh preparations (Figure 4.17) and of proteins prepared at least 12 months previously (Figure 4.18), and the results are summarised in Figure 4.19.

It is unlikely, however, that any of the additional disulphide bonds formed in the presence of 50mM Cu^{2+} are relevent to the disulphide bond pattern found *in vivo*, as the proteins may be at least partially denatured under these high [Cu^{2+}] conditions. The evidence for this is three-fold.

Firstly, the older samples that had been frozen and thawed more often showed a higher degree of additional disulphide bond formation than the fresh samples. Secondly, examination with the electron microscope of the older samples after incubation with 50mM Cu^{2+}, proved difficult as very few complete particles were present, and in several samples no particles were detected (D. Notman, personal communication). Similarly, fresh preparations of full-length mutant proteins after incubation with 50mM Cu^{2+} were particulate but many were incomplete or appeared "broken", and the truncated particles had dissociated and could not be seen. In all

Figure 4.17

SDS-Polyacrylamide Gel Electrophoresis of Freshly-Prepared Proteins following Incubation with Cu^{2+}

Approximately 5μg of all proteins were incubated for 20 minutes with either

a: 50μM Cu^{2+}

or

b: 50mM Cu^{2+}

and immediately electrophoresed through a 15% SDS-Polyacrylamide gel.

m = monomer

d = dimer

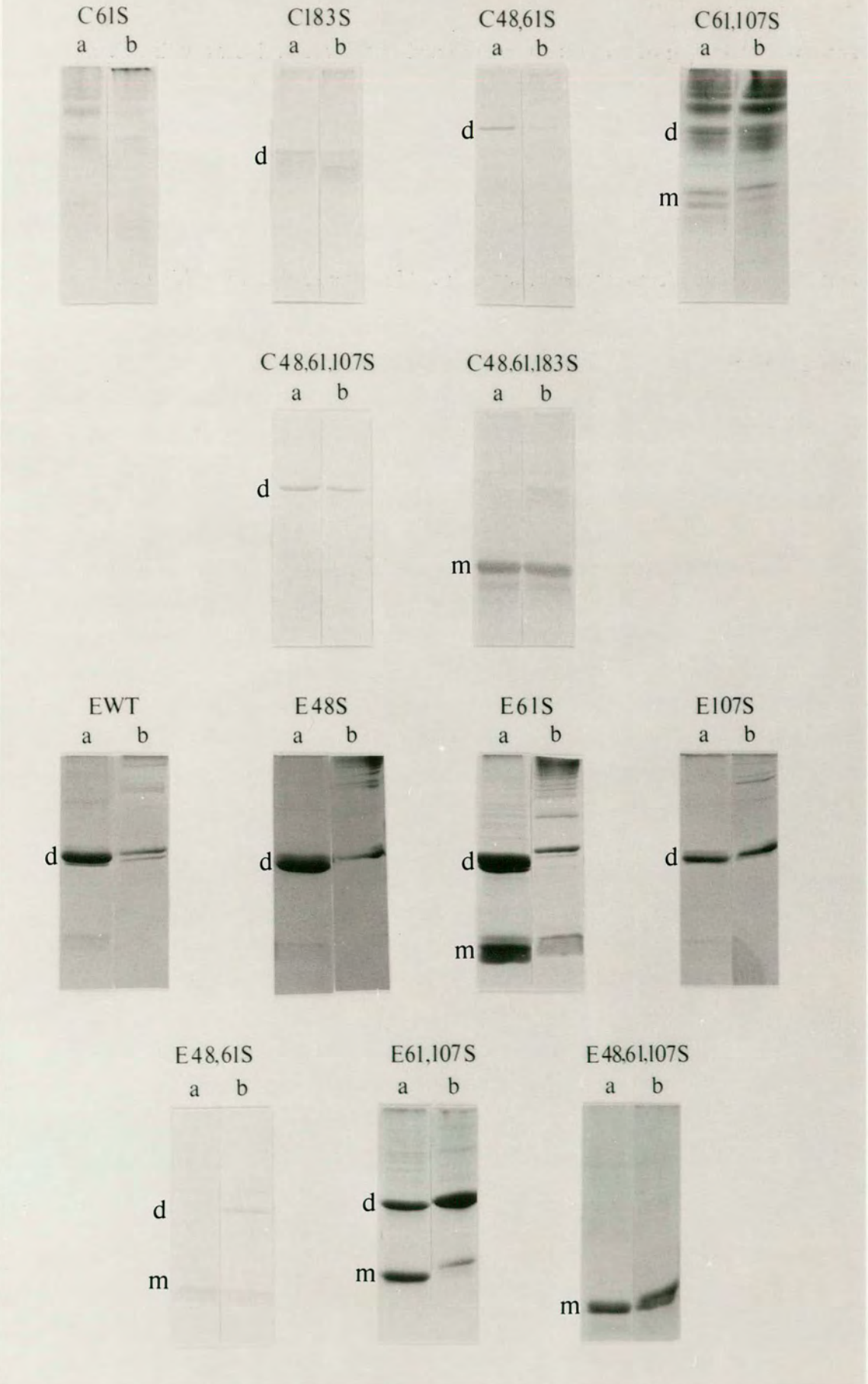

C61S
a b
C183S
a b
d
C48,61S
a b
d
C61,107S
a b
d
m
C48,61,107S
a b
d
C48,61,183S
a b
m
EWT
a b
d
E48S
a b
d
E61S
a b
d
m
E107S
a b
d
E48,61S
a b
d
m
E61,107S
a b
d
m
E48,61,107S
a b
m

Figure 4.18

SDS-Polyacrylamide Gel Electrophoresis of Older (Air-Oxidised) Proteins following Incubation with Cu^{2+}

Approximately 5μg of all proteins were incubated for 20 minutes with either

a: 50μM Cu^{2+}

or

b: 50mM Cu^{2+}

and immediately electrophoresed through a 15% SDS-Polyacrylamide gel.

C61,107S and E61,107S were not investigated as only fresh preparations of these proteins were available.

m = monomer

d = dimer

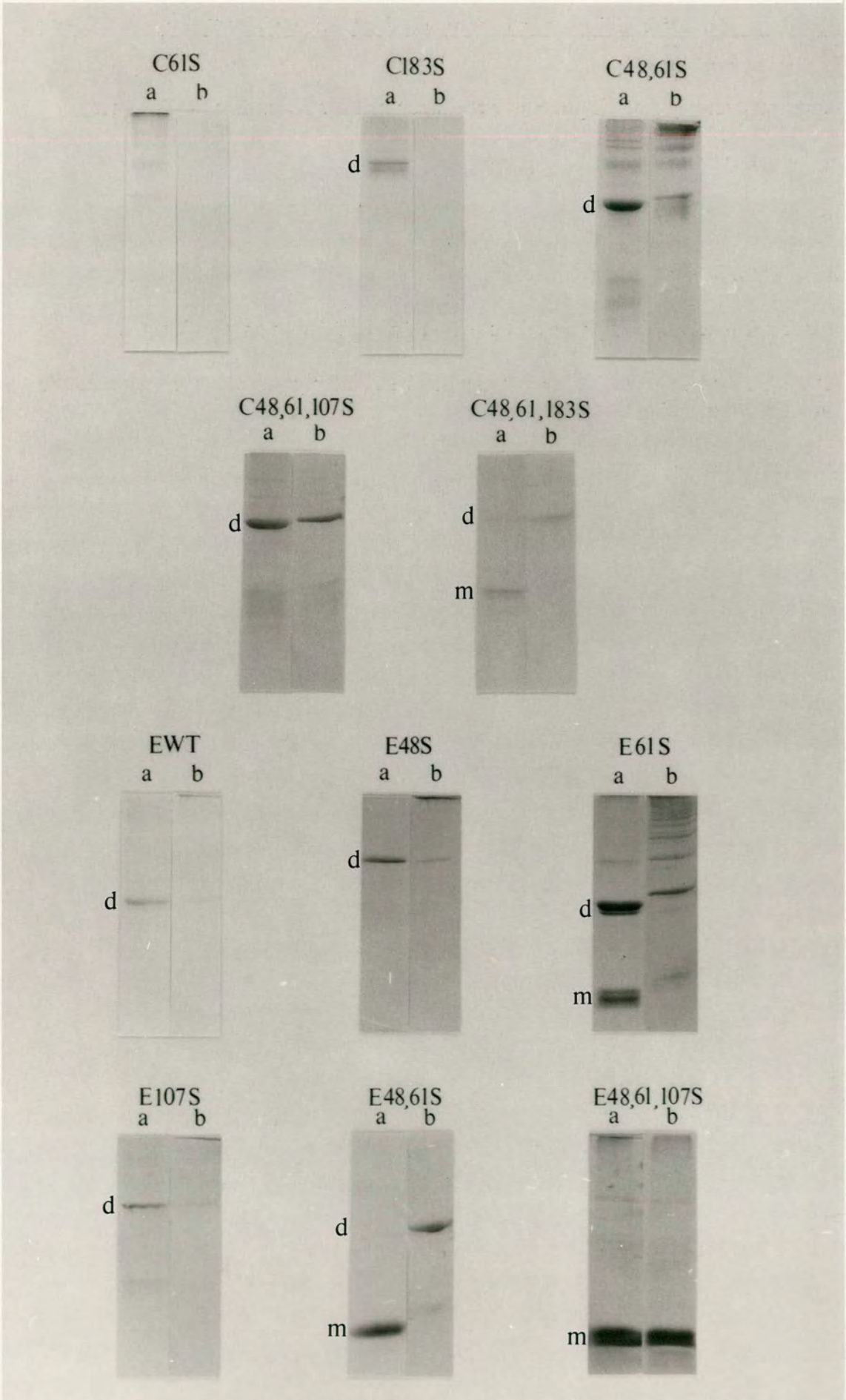

C61S
a b
C183S
a b
d
C48,61S
a b
d
C48,61,107S
a b
d
C48,61,183S
a b
d
m
EWT
a b
d
E48S
a b
d
E61S
a b
d
m
E107S
a b
d
E48,61S
a b
d
m
E48,61,107S
a b
m

Figure 4.19

Summary of Species Present in Gels through which proteins incubated with 50μM and 50mM Cu^{2+} were electrophoresed (Figures 4.17 and 4.18)

ND = no data, as only fresh preparations were available.

Protein	50μM Cu^{2+}	50mM Cu^{2+}	
		fresh	older
C61S	minor multimers	larger multimers	does not enter gel
C183S	100% dimer	dimer + larger	does not enter gel
C48,61S	100% dimer	dimer + larger	dimer + larger
C61,107S	minor multimers	fewer multimers	ND
C48,61,107S	100% dimer	100% dimer	100% dimer
C48,61,183S	100% monomer	monomer + dimer	100% dimer
EWT	100% dimer	dimer + larger	dimer + larger
E48S	100% dimer	dimer + larger	dimer + larger
E61S	monomer + dimer	monomer + dimer + larger	monomer + dimer + larger
E107S	100% dimer	dimer (+ larger)	dimer (+ larger)
E48,61S	100% monomer	monomer + dimer	100% dimer
E61,107S	monomer + dimer	increased dimer	ND
E48,61,107S	100% monomer	100% monomer	100% monomer

proteins examined, the majority of the protein ran in the gel. If, due to denaturation or some other reason, these multimers did not associate to form particles they would not be visible under the electron microscope but their mobility in the gel would not be affected.

Finally, as shown in Figure 4.20, the extra bonds formed under these conditions are most likely between two Cys107 amino acids, and the proteins in which only Cys48 would be available for additional disulphide bond formation form very few additional disulphide bonds. If it is therefore assumed that the majority of additional disulphide bonds are between two Cys107 residues, then Cys107-Cys107 bonds are formed between different HBcAg subunits in different mutants. It is unlikely that in its native conformation amino acid 107 would be in a position where it would be available to fulfil the stringent requirements for disulphide bond formation with either of two Cys107 residues on two alternative subunits. However, if HBcAg is a flexible protein, it may, very rarely, be in a conformation favourable for such disulphide bond formation; and the highly oxidising nature of the environment (50mM Cu^{2+}) may then cause the protein to be "locked" into this structure. However, it seems unlikely that a structural protein such as HBcAg which forms particles of consistent structure would be sufficiently flexible for this to occur.

The information obtained about the additional disulphide bonds formed under these conditions is useful, however, in terms of determination of the probability of the presence or absence of an internal disulphide bond. By the model deduced previously

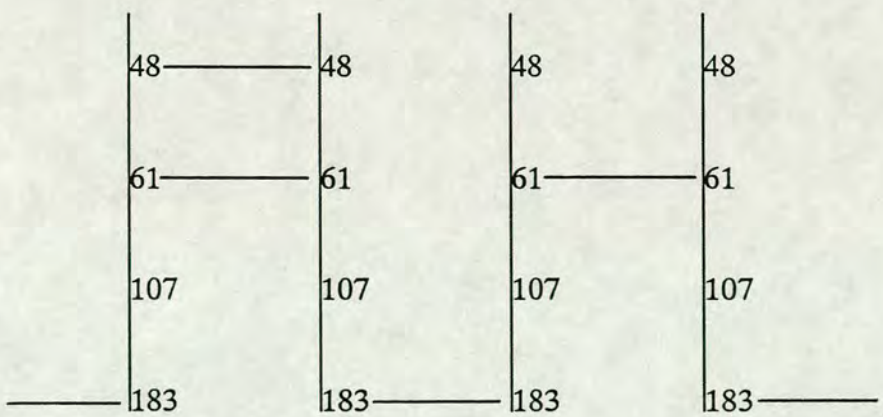
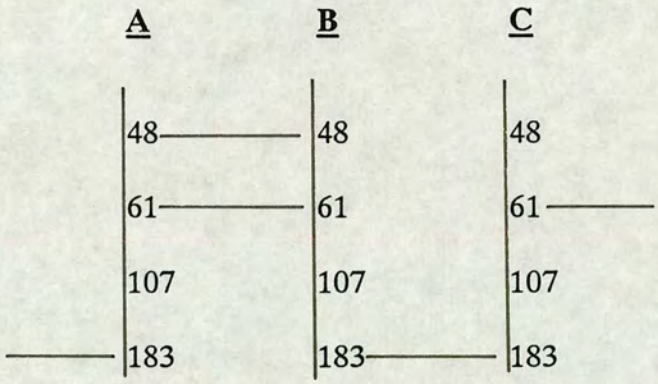


Figure 4.20

Position of Additional Disulphide Bonds Formed Following Incubation with 50mM Cu^{2+}

Samples C61,107S and E61,107S were freshly-prepared samples (Figure 4.17), indicated by *.

All other samples were older (air-oxidised) (Figure 4.18).

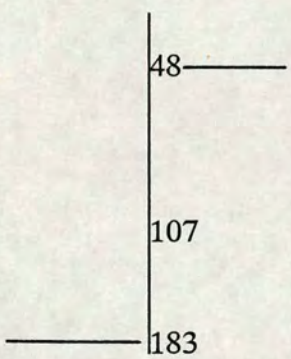


Protein	+ 50mM Cu^{2+}	Additional Bonds Between	
		Cysteines	Molecules
C61S	does not enter gel	107-107	A and B
C183S	does not enter gel	107-107	B and C
C48,61S	d + larger	107-107	A and B
C61,107S*	d + larger	48-48	A and B
C48,61,107S	100% d	no additional bonds	
C48,61,183S	100% d	107-107	either pair
EWT	d + larger	107-107	B and C
E48S	d + larger	107-107	B and C
E61S	m + d + larger	107-107	B and C
E107S	d + larger	48-48	B and C
E48,61S	100% d	107-107	either pair
E61,107S*	d + larger	48-48	either pair

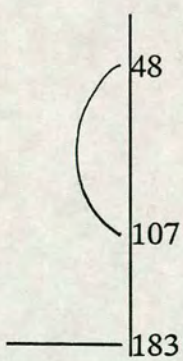
the only cysteines available to participate in an internal disulphide bond are Cys48 and Cys107, and Cys48 is available in only 50% of molecules, as in the remaining 50% of molecules it is involved in a Cys48-Cys48 bond. Therefore an internal disulphide bond could be present in only a maximum of 50% of molecules and could be between only Cys48 and Cys107. To determine whether this is in fact the case it is useful to predict the additional disulphide bond-forming potential of those proteins containing cysteines 48 and 107, if an internal Cys48-Cys107 bond was indeed present and then to compare this to the actual behaviour of the proteins after exposure to 50mM Cu^{2+} (Figure 4.18). Older protein preparations were examined as these formed a larger number of disulphide bonds after incubation with 50mM Cu^{2+} than the fresh preparations.

C61S

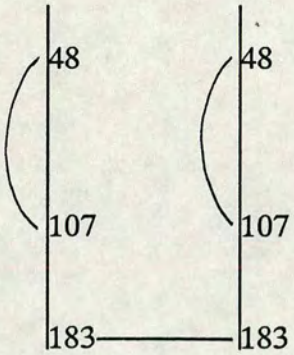
If 50% of Cys48 are present as



and 50% of Cys48 are present as



then the probability of

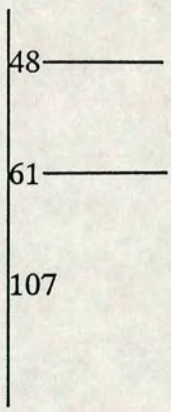


arising is 50% x 50% = 25%.

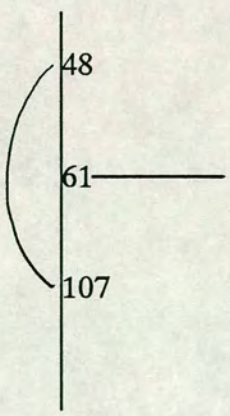
As no further bonds can be formed involving this dimer, it would be expected that 25% of the total protein on the gel would be dimer. This is not the case, however, as the protein does not enter the gel after incubation with 50mM Cu^{2+} (Figure 4.18), indicating the presence of larger multimers only, so that at least 80% of dimers are linked together by a bond invloving Cys48 and/or Cys107 and no dimers are present.

C183S

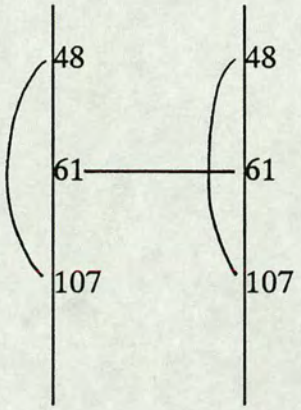
If 50% of Cys48 are present as



and 50% of Cys48 are present as



then the probability of

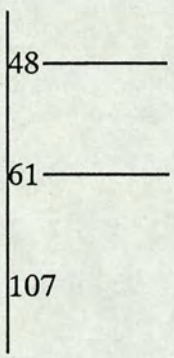


being formed is 50% x 50% = 25%.

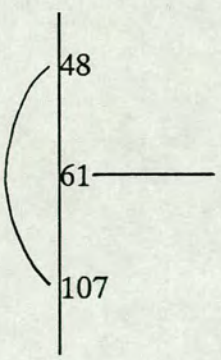
No further multimers can be formed by this dimer as there are no free cysteines and so 25% of the protein would be expected to run in the gel as dimers. As with C61S this is not the case and all the protein, after incubation with 50mM Cu^{2+}, exists as larger multimers which do not enter the gel and there are no dimers (Figure 4.18).

EWT

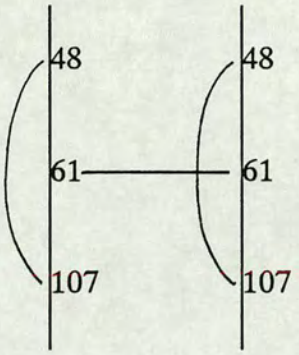
If 50% of Cys48 are present as



and 50% of Cys48 are present as



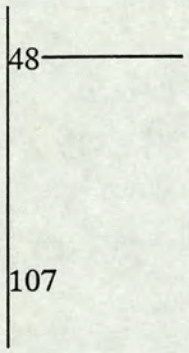
then the probability of



being formed is 50% x 50% = 25%. As this dimer can form no further disulphide bonds, having no free cysteines, it would be expected that 25% of the total protein would be dimer. However, examination of the EWT gel of older preparations (Figure 4.18) with a densitometer (Figure 4.12) demonstrates that only 7% of the total protein exists as dimer. That is, the density of the dimer band in the gel at 50mM Cu^{2+} is only 7% of the density of the total protein in the gel when no Cu^{2+} is present.

E61S

If 50% of Cys48 is present as



and 50% of Cys48 is present as

then as the latter has no free cysteines it can form no multimers, and so 50% of the

total protein would be expected to be in monomer form. However, densitometric scanning (Figure 4.12) of the E61S gel (Figure 4.18) shows that only a maximum of 5% of the total protein is monomer. As only protein that has entered the gel is measured, those multimers too large to enter the gel cannot be included in the calculation and so the figure obtained for monomer as a proportion of total protein is the maximum possible.

Therefore, in particles formed from all four proteins more multimers are formed than would be expected if an internal Cys48-Cys107 disulphide bond was present in 50% of cases. These results do not, however, rule out the possibility that an intra-molecular disulphide bond can be formed, but in the truncated proteins only, and in significantly less than 50% of cases (a maximum of 7% in the case of EWT).

4.3 Discussion

The nucleocapsid of HBV fulfils several roles *in vivo*. Formation of complete, replication-competent nucleocapsids requires the initial interaction of HBcAg with both the packaging signal on the RNA pre-genome (Junker-Niepmann *et al.*, 1990) and the viral polymerase (Bartenschlager *et al.*, 1990; Hirsch *et al.*, 1990). As it has been demonstrated that HBcAg alone can form nucleocapsid-like structures, both in *E. coli* (Stahl *et al.*, 1982; Cohen and Richmond, 1982) and other heterologous hosts (Kniskern *et al.*, 1986; Roosinck *et al.*, 1986; Nassal, 1992a)), the interactions described above are not required for capsid assembly. Once the complete nucleocapsid is formed, it is thought to play an active role in viral replication. Specifically, the COOH-terminus of HBcAg is essential for viral genome replication and may be involved in binding, at least temporarily, to the nascent viral DNA (Nassal, 1992b). In addition, envelopment of the nucleocapsid requires its interaction with membrane-embedded HBsAg and so the nucleocapsid must be in the correct configuration to facilitate this.

Therefore, the nucleocapsid must maintain its structure, both internally and externally and throughout the viral life cycle, for all these roles to be fulfilled successfully. One means by which this structure could be maintained is by the stabilising effect of disulphide bonds. HBcAg has four cysteines which are completely conserved among mammalian hepadnaviruses and all four are therefore likely candidates for involvement in disulphide bonds. In this work the role of each was assessed by the effect of its elimination, singly and in combination with others, on particle structure and particle stability.

Firstly, the effect on particle structure of substituting cysteine residues with serine residues was determined. High molecular weight particles were shown to be formed from all 14 mutant proteins by their behaviour during purification. That these particles have HBc/HBe antigenicity comparable to that of wild-type particles was determined by radioimmune assay, and that these particles were of normal morphology was further confirmed by electron microscopic examination. This confirms the observations of others (Gallina *et al.*, 1989; Stahl and Murray, 1989; Salfeld *et al.*, 1989; Birnbaum and Nassal, 1990; Nassal, 1992b) that HBcAg-derived proteins lacking the arginine-rich carboxy-terminus form particles of morphology indistinguishable from those formed by wild-type full-length proteins. The slightly increased size of particles formed from truncated proteins which has been reported (Gallina *et al.*, 1989) was not observed with either the wild-type truncated or mutant truncated protein particles produced in this work, in agreement with the observations of Zheng *et al.* (1992). Thus, the main structural domain of HBcAg does not include, or does not depend upon, the arginine-rich C-terminus which is not involved in essential interactions with other HBcAg molecules to form particles. The C-terminus is thought to be (at least largely) internal to the particle (Zhou *et al.*, 1992), and may be thought of as a separate domain of the protein, possibly protruding into the interior of the capsid. The nucleocapsid of Tomato Bushy Stunt Virus also consists of 180 subunits and has icosahedral symmetry. In a similar manner to Hepatitis B Virus core particles, these capsids are composed of only one monomer type, which also has an internal domain, in this case the amino terminal region of the protein. X-ray crystallography has

demonstrated that the internal domains of these monomers interdigitate to form an internal shell (Harrison, 1985). It is tempting to speculate that this may also be true for HBV nucleocapsids and that this internal shell is stabilised by Cys183-Cys183 disulphide bonds.

These results also demonstrate that substitution of cysteines with serines causes no gross alteration to particle structure as neither the ability to form particles nor the reaction of these particles with polyclonal anti-HBc/anti-HBe antibodies is affected. The radioimmune assays performed do not exclude the possibility that subtle structural changes have taken place, perhaps exposing HBe epitopes rather than HBc epitopes, or *vice versa*, but these have had no effect on the ability of the proteins to form particles. However, it has been demonstrated that proteins also produced in *E. coli* and almost identical to those produced in this work, (with the exception that 5 additional amino acids (MITNS) are present before amino acid 1 of HBcAg, the truncated proteins terminate at amino acid 144, and cysteines are mutated to serines or alanines) react to the same extent as wild-type full-length HBcAg with a panel of antisera raised against synthetic peptides corresponding to overlapping regions of HBcAg (Zheng *et al.*, 1992). In addition, another panel of mutant proteins produced in *E. coli* which are very similar to those studied in this work (with the exception that no non-HBcAg sequences are present and truncated proteins terminate at amino acid 149) react with a monoclonal antibody which "preferentially recognises particulate core protein" (Nassal *et al.*, 1992). This strongly suggests that such alterations to structure have not in fact occurred.

Therefore the cysteines of HBcAg are not required for its correct folding. In addition they are not required for particle formation as all mutant proteins form "normal" particles even if only one cysteine is present (in the context of the full-length protein) or if no cysteines at all are present (in the context of the truncated protein). However, as may be expected, the mutant proteins formed particles with stability proportional to the number of cysteines present, with the exception of cysteine 107 (unpublished observations). This was true in terms of denaturation due to both repeated

freeze/thawing of preparations and denaturation due to more artificial (50mM Cu^{2+}) conditions.

In summary, conclusions derived from the mutant proteins should be able to be extrapolated to the native protein as the protein structure does not appear to be altered by mutation of cysteines to serines.

This should also be true for proteins treated with 50μM Cu^{2+}, the concentration at which all "native" disulphide bonds have formed. While there is no direct evidence presented here that this is the case, there are three supporting lines of evidence. Firstly, as the disulphide bonds formed after exposure to 50μM Cu^{2+} were essentially the same as those formed after extended air oxidation, it is reasonable to conclude that both are in the same physical state. Secondly, the protein preparations used in the radioimmune assay and the particles examined under the electron microscope were at least 6 months old at the time and were shown to have the same proportion of disulphide bonds as both fresh, and equivalently aged, preparations that had been exposed to 50μM Cu^{2+}. The structure of these samples was normal and, if these samples are indeed in the same state as the Cu^{2+}-exposed particles then they are unaltered by 50μM Cu^{2+}. Thirdly, this is supported by the observation of Nassal *et al.* (1992) that proteins almost identical to those studied here (but containing only HBcAg-derived amino acids and terminating at amino acid 149 in the case of truncated proteins), after incubation with up to 100μM Cu^{2+}, were stable particles, as determined by sucrose gradient analysis and ELISA assay using a monoclonal antibody that preferentially recognised particulate HBcAg. This contrasts sharply with the behaviour of these particles after incubation with 2mM Cu^{2+} which caused truncated particles, except those lacking Cys48, curiously, to dissociate (Nassal *et al.*, 1992). Similarly, examination under the electron microscope of samples (this work) that had been incubated with 50mM Cu^{2+} demonstrated that particles formed from full-length proteins were now particulate but rarely intact and often "broken" in appearance, and that truncated particles had dissociated to an extent whereby they could no longer be detected. It seems likely, therefore, that while high (mM)

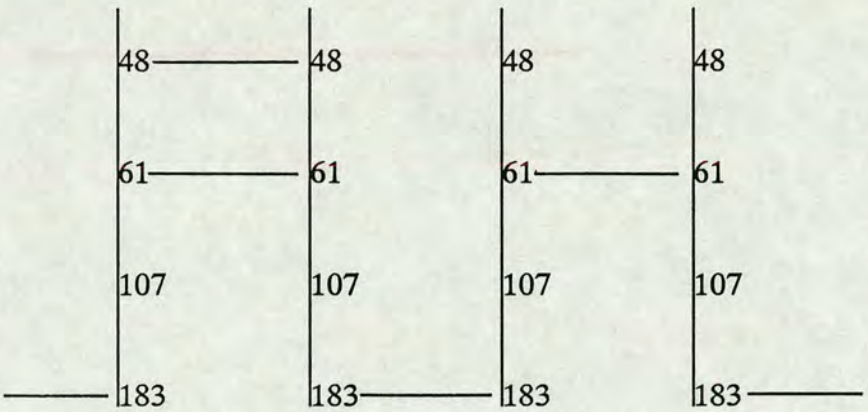
concentrations of Cu^{2+} cause deleterious structural effects the low concentration of 50μM causes no significant such effects.

Inter-molecular Disulphide Bonds

The results produced by polyacrylamide gel electrophoresis of the panel of mutant proteins under non-reducing conditions clearly define the inter-molecular disulphide bonds present within core particles. As the maximum size of disulphide-linked species produced by the wild-type truncated protein is dimer this allowed more confident allocation of disulphide bond participants than with the full-length proteins, as in the former case all the protein entered the gel and could therefore be quantified readily. In contrast, of the full-length protein that did not enter the gel, it could not be said that it was a 100% disulphide-linked lattice, but only that it existed as disulphide-linked multimers of at least approximately 10 subunits (approximately 220kD) which would require a minimum of only 80% of potential disulphide bonds to have formed. Another conclusion that could be immediately drawn from the dimeric nature of wild-type truncated protein is that any disulphide bonds involving Cys48, 61 or 107 are between the same subunits.

Full-length and truncated proteins that contain only Cys107 contain no disulphide bonds in their particles, and the presence of Cys107 does not affect the disulphide bond pattern of any of the mutant protein particles. Thus, Cys107 was excluded from involvement in inter-molecular disulphide bonds. Also, in all cases the behaviour of the full-length proteins mirrored that of their truncated counterparts and confirmed the existence of a Cys183-Cys183 bond in 100% of cases as those truncated mutants that were monomeric became dimers upon addition of Cys183 and those that had been dimers no longer entered the gel. Cys61 is also involved in a disulphide bond with Cys61 on an adjacent molecule in 100% of cases, but between monomer pairs not linked by Cys183-Cys183 bonds. Curiously, Cys48 exists as a disulphide bond participant in only 50% of cases, but always with Cys48 on the same molecule as that to which the Cys61 is bonded. Therefore the proposed model for inter-molecular

disulphide bond arrangement is



Intra-molecular Disulphide Bond

Nassal *et al.* (1992) reported that their mutant protein equivalent to the E61S protein of the work described here existed initially as monomer only, 50% of which was converted to dimer upon exposure to 50μM Cu^{2+}, in agreement with this work. However, the monomer ran in a non-reducing polyacrylamide gel as two close bands, which could be converted to only the slower-migrating species upon addition of an excess of DTT. It was therefore concluded by the authors that an internal disulphide bond was present in approximately 30% of the monomers, causing it to migrate faster in the gel. This explanation was further supported by the observation that only the slower-migrating form was converted into dimer upon exposure to Cu^{2+}. However, this was not observed in any other mutants by Nassal *et al.* and has not been observed with any of the proteins, including E61S, in the work presented in this thesis. It is noteworthy, however, that some proteins examined in this work do appear occasionally as a doublet in the gel, both when non-reduced (including C183S, Figure 4.14 A and D; E61S, Figure 4.15B; C61,107S, Figure 4.17; EWT, Figure 4.17; C183S, Figure 4.18; E61S, Figure 4.18) and reduced (not shown). However, this was not observed consistently and does not correlate with the presence or absence of Cys48 or Cys107 and a doublet is also observed for dimers where the presence of an internal disulphide bond would have precluded dimer formation (including E61S, Figures 4.15B and 4.18). A similar phenomenon was also observed by Zheng *et al.* (1992) with their

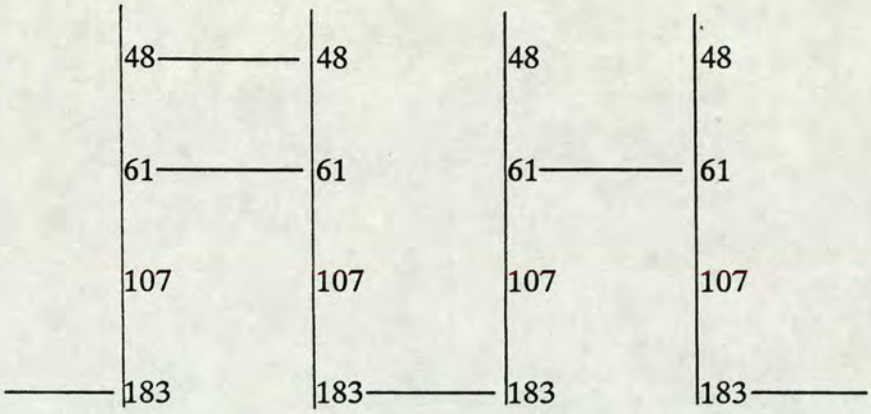
mutant proteins equivalent to C61S, C48,61S and C61,107S. It is possible that limited proteolysis has occurred in both cases, despite the presence of protease inhibitors, but as this was not observed consistently with specific mutants in this work it is unlikely to be due to any structural differences between different mutants.

The question of the presence of an internal disulphide bond was answered indirectly in this work. As has been discussed above, exposure to 50mM Cu^{2+} resulted in disruption or dissocation of particles, but this provided further information concerning the availability of cysteines known not to be involved in inter-molecular disulphide bonds, as only those cysteines not already involved in disulphide bonds would be available to form the new disulphide bonds forced by 50mM Cu^{2+}. It is unlikely that 50mM Cu^{2+} causes breakage of existing disulphide bonds as this would require the presence of an excess of reducing agent and not an oxidising agent such as Cu^{2+} and so the novel protein forms seen are unlikely to have arisen *via* disulphide bond rearrangement. In addition, if disulphide bond breakage was taking place, the same disulphide bond pattern may be expected for all mutant proteins after exposure to 50mM Cu^{2+}, which is not in fact the case. As EDTA was added after incubation with Cu^{2+} and before electrophoresis, it is unlikely that the Cu^{2+} ions are an integral part of the disulphide bonds formed after exposure to 50mM Cu^{2+} and are more likely merely catalysts for the reaction.

In the proteins in which Cys48 and Cys107 (necessarily the only cysteines between which an intra-molecular disulphide bond could form) are both present, their behaviour was inconsistent with the presence of an intra-molecular Cys48-Cys107 bond. Proteins in which 25% would be expected to exist as dimer, or 50% as monomer, did exist as monomer and dimer, respectively, in the case of the truncated proteins only, but in a much lower percentage than predicted. This may indicate the presence of an internal disulphide bond in a small number of cases but this is unlikely. Such a model would require Cys48 to exist in at least 3 forms - as SH, as inter-molecular disulphide bond and as intra-molecular disulphide bond, which would in turn require three different forms of HBcAg monomer to exist.

There is precedent for the presence of three quasi-equivalent forms of monomer in the nucleocapsids of Tomato Bushy Stunt Virus and Turnip Crinkle Virus which are also composed of 180 subunits, with icosahedral symmetry (Harrison, 1985). This symmetry is known as T=3 symmetry which means that the 60 subunits of the basic structure each consist of 3 further subunits, hence T=3. Crystallographic studies of these nucleocapsids indicate that the 3 sub-subunits exist in 3 quasi-equivalent forms in order to accommodate 180 total subunits into the overall structure, but this means that 33% of monomers are in each form and that very small proportions of subunits (e.g. the 7% for EWT) would not fit with this model precisely. It is more likely that denaturation is not complete and that the full disulphide bonding capacity of the denatured protein has not been attained, and that the monomers and dimers present contain free sulphhydryl groups. This could be investigated by denaturing the protein completely, perhaps by heating, prior to incubation with Cu^{2+}. It is therefore unlikely that Cys48 and Cys107 are involved in an intra-molecular disulphide bond in the native particle. This was also the conclusion of Zheng *et al.* (1992) who determined, by means of reaction of their panel of mutant proteins with Ellman's Reagent, the number of free SH groups per HBcAg molecule to be 0.5 when Cys48 was present and Cys107 was absent; 1.0 when Cys107 was present and Cys48 was absent; and 1.5 when both Cys48 and Cys107 were present. Their work also indicated that Cys107 was buried within the particle as it reacted with Ellmann reagent only after exposure of the particles to 0.1% SDS. This denaturation may be equivalent to the denaturation of the protein caused by 50mM Cu^{2+} in this work which allowed the free Cys 107 to become available for disulphide bond formation.

Thus the overall model for the disulphide bond arrangement of HBcAg is:



This model is in complete agreement with that proposed by Zheng *et al.* (1992) and Nassal *et al.* (1992). In addition, it is agreed that neither cysteines nor disulphide bonds are required for formation of normal core particles in *E. coli.* This has now been taken one stage further by Nassal (1992) who transiently transfected human hepatoma-derived HuH7 cells with HBV genomic DNA in which all 4 cysteine codons of the full-length C gene had been mutated to serine codons. Comparison of the complete virions produced after transfection with this mutant, with those produced after transfection with wild-type DNA, showed that complete virions were indeed produced. Therefore, lack of cysteines does not critically impair any stages in the viral life cycle subsequent to virus infection. However, minor differences between wild-type and mutant virus particles were noted. "Substantially fewer" enveloped cores were observed with the mutant than with wild-type but this may reflect the reduced stability of mutant virions rather than reduced efficiency of envelopment. The former is thought likely as those mutant virions formed were intact and, in contrast to wild-type, isopycnic centrifugation during purification of these virions resulted in viral nucleic acid being present along with HBcAg in a position characteristic of naked cores. This would suggest that a proportion of the virions had lost their envelopes during centrifugation, in contrast to the wild-type virions which showed no nucleic acid at this position. Data for this observation was not shown and no indication was given of the relative quantity of naked, nucleic acid-containing cores present relative to intact virions. Therefore, cysteines may be involved not only in the stability of nucleocapsids but also in the stability of virions themselves by enhancing the stability of nucleocapsid/envelope interactions. As Cys183 is thought to be in the interior of the

core particle and Cys107 is not accessible in the absence of denaturing agent, Cys48 and Cys61 are more likely to be available for involvement in direct Core/Surface interactions. Cys61 is involved in disulphide bonds between HBcAg molecules, but 50% of Cys48 is free. Zheng *et al.* (1992) have shown that Cys48 readily reacts with sulphhydryl reagents even in the absence of denaturing agents and so is exposed in the native particle. Antibodies produced against synthetic peptides derived from the Cys48 region of the protein also bind native full-length and truncated particles, giving further credence to the proposition that Cys48 may be on or near the surface of the nucleocapsid (Zheng *et al.*, 1992). As the surface protein HBsAg has 14 cysteines it can be envisaged that the 50% of Cys48 not involved in an inter-molecular disulphide bond with another HBcAg molecule may be involved in an inter-molecular disulphide bond with a cysteine in a HBsAg molecule. As nucleocapsids lacking Cys48 are still enveloped, the role of a Cys48-HBsAg disulphide bond may be similar to the role of the disulphide bonds of the nucleocapsid in stabilising rather than determining inter-molecular interactions.

Another difference in the behaviour of wild-type and mutant nucleocapsids within complete virions noted by Nassal (1992b) was a consistently lower proportion (15% as opposed to 35%) of relaxed circular (RC) DNA to linear DNA in the mutant. Therefore, formation of RC DNA from linear DNA, which involves the primer transfer reaction, may be less efficient in the absence of the cysteines of HBcAg. In particular, Cys183 may be important for this as it is thought to lie within the capsid and at the COOH-terminus of the region of HBcAg proposed to bind the nascent HBV DNA formed by reverse transcription. It would therefore be interesting to investigate the behaviour of proteins lacking only Cysteine 183 in this regard.

The role of the cysteines within HBcAg remains unclear. Their strict conservation suggests a significant role, but their elimination has no apparent effect on core particle formation, at least *in vitro*. Their observed role in conferring stability to core particles and to virions may be important but equally may not be required under conditions of normal infection where exposure to environments and treatments as harsh as those

imposed *in vitro* may not occur. As woodchuck hepatitis virus core antigen displayed properties identical to HBcAg with regard to cysteines (Zheng *et al.*, 1992) the woodchuck model would be an appropriate one in which to test the behaviour of mutants lacking cysteines during the normal course of infection. This would also allow investigation of the role of the cysteines in initial infection of cells, which cannot be tested by transfection of tissue culture cells with DNA only. Alternatively, this could be achieved with HBV in the HepG2 system described by Bchini *et al.*, 1990.

A further possibility is that the cysteines are required for the normal function of HBeAg. The mature HBeAg contains cysteines 48, 61 and 107 and an additional cysteine within the N-terminal 10 amino acids of the pre-core region which remain attached to amino acid 1 of HBcAg after removal of the signal sequence. This additional cysteine is thought to be important in determining the disulphide bond content and structure of HBeAg as when it is present monomers only and no particles are observed, while when it is substituted particles formed from disulphide-linked dimers are produced (Schlicht and Wasnauer, 1991). It has been proposed that this additional cysteine may form an intra-molecular disulphide bond, possibly with Cys107, thereby altering the overall structure of the dimers and preventing particle formation. However, until the precise role of HBeAg during HBV infection is known, the effect of removal of cysteine residues cannot be determined.

Finally, the conservation of the cysteine residues in HBcAg may be telling us that although mutation to serine does not affect the ability of the polypeptides to assemble into particles that can function apparently normally, the slight change in total free energy of these particles in relation to that of particles formed from the wild-type protein may be disastrous in the competitive world of evolutionary selection.

Chapter 5: The Role of the Carboxy-Terminal Region of HBcAg in Determining Disulphide Bond Formation

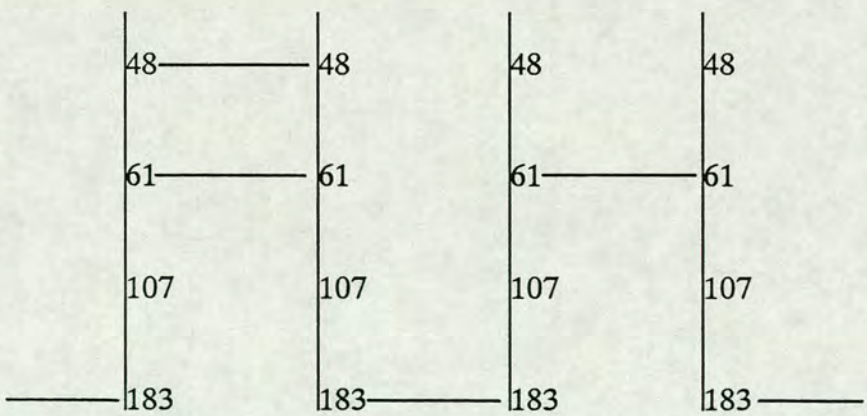
5.1 Introduction

The unexpected behaviour of the full-length mutant C183S in terms of the lack of Cys48-Cys48 and/or Cys61-Cys61 disulphide bond formation in the absence of extended air oxidation or Cu^{2+}-catalysed oxidation, unlike the other proteins which formed the majority of these bonds shortly after liberation of the protein from the bacterium, suggested that the nature of the carboxy-terminus may influence determination of disulphide bond formation in the remainder of the protein. This was investigated by the examination of the behaviour of both full-length and truncated core proteins, and fusion proteins constructed previously in our laboratory (Stahl and Murray, 1989) which differ in the nature of their carboxy terminal regions. Their behaviour was monitored in terms of the ability of the proteins to form particles and the mobility of the proteins in non-reducing polyacrylamide gels.

5.2 Results

As shown previously, freshly-prepared C183S ran in the gel as 100% monomer and no disulphide bonds were present (Figure 4.13).

In contrast, freshly-prepared CWT, which is the same length as C183S but has a cysteine at its carboxy terminus in place of the serine of C183S did not enter the gel (Figure 4.13) indicating that, by the model



at least 80% of the possible Cys48-Cys48 and/or Cys61-Cys61 bonds were present,

thus causing the formation of multimers too large to enter the gel. Such bonds were completely absent in freshly-prepared C183S.

Similarly, the freshly-prepared truncated protein EWT which contains the same cysteines as C183S but lacks amino acids 149-183, runs in the gel as dimer (15%) and monomer (85%) (Figure 4.15). In order to form a dimer, by the model deduced previously (above) a Cys61-Cys61 and/or Cys48-Cys48 bond must be present between 15% of molecules. Such bonds were completely absent in freshly-prepared C183S.

Therefore, either the presence of cysteine 183 or the absence of the arginine-rich region of the protein allow immediate formation of bonds that were formed in C183S only after prolonged air oxidation, or after oxidation catalysed by Cu^{2+} ions.

Further investigation of this phenomenon was achieved by utilisation of core fusion proteins produced previously in our laboratory (Stahl and Murray, 1989) which are described below:

HBcpreS1(1-20)

$_{1}$[β-gal]$_{8}$ EFH $_{3}$[HBcAg]$_{144}$ ISSRELGYS $_{1}$[preS1]$_{20}$ LDRCP

Total amino acids = 187
Total cysteines = 4

This protein contains cysteines 48, 61 and 107 of HBcAg and a cysteine as the penultimate amino acid of the protein (Figure 5.1A).

HBcpreS1(1-36)

$_{1}$[β-gal]$_{8}$ EFH $_{3}$[HBcAg]$_{144}$ ISSRELGYS $_{1}$[preS1]$_{36}$

Total amino acids = 198

Total cysteines = 3

This protein contains only cysteines 48, 61 and 107 of HBcAg. It is a similar length to the wild-type HBcAg produced in this work (192 amino acids) but the arginine-rich carboxy-terminus of the full-length HBcAg has been replaced by a sequence unremarkable in terms of amino acid content (for these purposes) or charge (Figure 5.1B).

HBcE46

$_{1}$[β-gal]$_{8}$ EFH $_{3}$[HBcAg]$_{144}$ IS $_{728}$[HIV env]$_{751}$

Total amino acids = 179

Total cysteines = 3

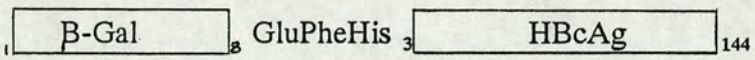
This protein contains only cysteines 48, 61 and 107 of HBcAg. It is shorter than the full-length wild-type HBcAg produced in this work, but longer than the truncated proteins (157 amino acids) (Figure 5.1C).

HBcpreS1(1-20) had previously been shown, by electron microscopy, to form particles of normal morphology (Stahl and Murray, 1989). Electron microscopic examination of HBcpreS1(1-36) and HBcE46 in this work reveals that they also form such particles (Figure 5.2). Immunological assays also demonstrated that the additional amino acids in all three fusion proteins are displayed on the surface of the particles. (Stahl and Murray, 1989). All fusion proteins used in this work had been prepared (by S. Bruce) at least 12 months before being assayed and so can be considered to have undergone

Figure 5.1

Amino Acid Sequences of Fusion Proteins

All three fusion proteins consist of amino acids:



followed by:

A. HBcpreS1(1-20)

Ile Ser Ser Arg Glu Leu Gly Tyr Ser [1]Met Gly Gln Asn Leu Ser Thr Ser Asn Pro Leu Gly Phe Phe Pro Asp His Gln Leu Asp[20] Leu Asp Arg **Cys** Pro

B. HBcpreS1(1-36)

Ile Ser Ser Arg Glu Leu Gly Tyr Ser [1]Met Gly Gln Asn Leu Ser Thr Ser Asn Pro Leu Gly Phe Phe Pro Asp His Gln Leu Asp Pro Ala Phe Arg Ala Asn Thr Asn Asn Pro Asp Trp Asp Phe Asn Pro[36]

C. HBcE46

Ile Ser Leu Pro Ile Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp Arg Asp Arg

Figure 5.2

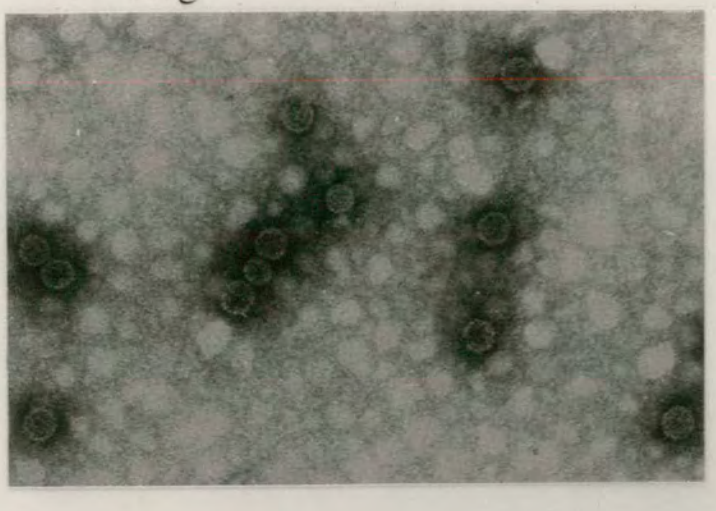
Electron Micrographs of Fusion Proteins

All particles are shown at a magnification of 124 000x and all microscopy was performed by P. Highton.

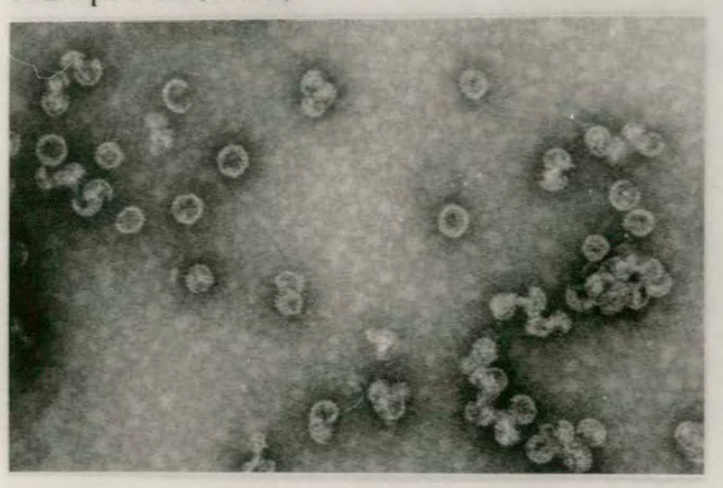
HBcpreS1(1-20) had been demonstrated previously by electron microscopy to form particles morphologically indistinguishable from those formed by wild-type HBcAg (Stahl and Murray, 1989).

wtHBcAg = TacpCore

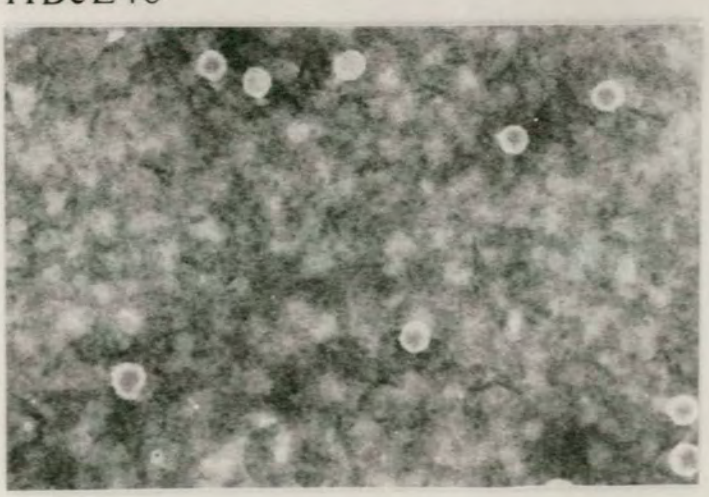
wtHBcAg



HBcpre S1(1-36)



HBcE46



extended air oxidation.

All fusion proteins were incubated at the appropriate temperature for 10 minutes in the absence of reducing agent and immediately electrophoresed through a 15% SDS-polyacrylamide gel. The same result was achieved when proteins were incubated at 0, 20, 40, 60, 80 and 100°C prior to electrophoresis and a representative Coomassie-stained gel is shown in Figure 5.3.

HBcpreS1(1-20), which contains 4 cysteines, behaves in a manner analagous to CWT in not entering the gel. In contrast, HBcpreS1(1-36) which is of similar size to CWT but has a different COOH-terminal region containing no cysteine at its terminus, behaves in a manner analagous to freshly-prepared C183S (which also has only cysteines 48, 61 and 107) in running in the gel only as monomer. HBcE46, which has only three cysteines and a shorter fusion region, behaves in a manner analagous to the freshly-prepared wild-type truncated protein (EWT) in running in the gel as both monomer and dimer. These results are summarised in Figure 5.4. There were insufficient amounts of the fusion proteins available to be able to examine the samples after incubation with Cu^{2+}, but as the samples had been prepared at least 12 months previously it would be expected that disulphide bond formation had proceeded significantly towards completion. (Full-length and truncated proteins, as described earlier, showed very similar disulphide bond complements after extended air oxidation and after incubation with Cu^{2+}. Sections 4.2.7)

5.3 Discussion

From the behaviour of the full-length, truncated and fusion proteins in non-reducing SDS-PAGE, it is apparent that the nature of the carboxy-terminus of HBcAg, while dispensable for particle formation, influences the ease of disulphide bond formation, or even the ultimate disulphide bond complement of the particle.

Figure 5.3

Non-reducing SDS-Polyacrylamide gel electrophoresis of mutant and fusion proteins

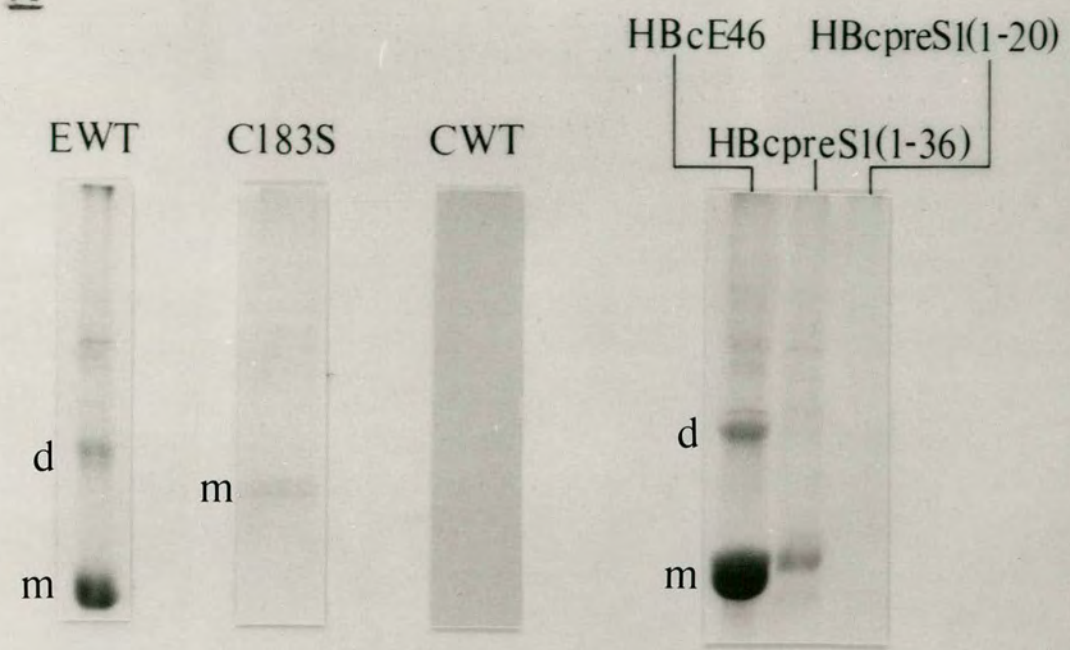
A. 5μg of all proteins were incubated at 100°C for 5 minutes in the absence of reducing agent and immediately electrophoresed through a 15% SDS-polyacrylamide gel which was subsequently stained with Coomassie Blue.

B. Gels were dried between two sheets of cellophane membrane and analysed with a densitometer. The absorbance at 550nm was measured along each gel track and the areas under the peaks of the resulting graphs measured by the densitometer. These figures were then converted to % monomer and % dimer.

m = monomer

d = dimer

A



B

	Densitometry	
EWT	85% monomer	15% dimer
C183S	100% monomer	
CWT	does not enter gel	
HBcE46	94 % monomer	6 %dimer
HBcpreS1(1-36)	100 % monomer	
HBcpreS1(1-20)	does not enter gel	

Figure 5.4

Summary of behaviour of mutant and fusion proteins during non-reducing SDS-polyacrylamide gel electrophoresis

Protein	Cysteines Present				Behaviour in Gel
C183S	48	61	107		100% monomer
CWT	48	61	107	183	does not enter gel
EWT	48	61	107		monomer and dimer
HBcpreS1(1-20)	48	61	107	C	does not enter gel
HBcpreS1(1-36)	48	61	107		100% monomer
HBcE46	48	61	107		monomer and dimer

While full-length and truncated proteins all formed the predicted disulphide bonds (Cys48-Cys48 and/or Cys61-Cys61) after extended air oxidation or Cu^{2+}-catalysed oxidation, only C183S had an absolute requirement for such oxidation as the other proteins all formed these bonds within a short time (less than five days) after liberation from the bacterium. As all proteins were in the same environment, with the same redox potential, the difference in ease of disulphide bond formation must be due to a difference in the conformation of the protein. However, the conformational difference must be only slight, as the disulphide bond(s) do ultimately form and this would not occur if the cysteine residues were kept permanently apart by the folded protein conformation. Where the COOH tail is absent (EWT) the bonds form immediately, and where the COOH- terminal is present but Cys183 is also present the bonds form immediately. It could therefore be imagined that the COOH-tail prevents the close contact required between Cys48-Cys48 and Cys61-Cys61 for disulphide bonds to form, but that if two COOH-tails are linked by a disulphide bond, such close contact is not prevented. However, while in no cases did Cys48-Cys48 and/or Cys61-Cys61 bonds form in C183S particles immediately, only approximately 15% of these potential disulphide bonds were formed in EWT immediately and it cannot be predicted how many such bonds were formed in CWT as it can be said only that all molecules were in the form of multimers of at least 10 molecules (which do not enter the gel). However, this would require a minimum of 80% of Cys48-Cys48 or Cys61-Cys61 bonds to have formed. Therefore, the presence of Cys183 may be more important in promoting tight/correct structure formation than the absence of the entire COOH tail. The function of the inner shell of the Tomato Bushy Stunt Virus capsid, formed from the internal domains of the component monomers, is thought to be to ensure that the viral shell closes around on itself correctly during assembly (Harrison, 1985). If the HBV core particles have a similar inner shell, its function may also be similar. Removal of the carboxy-terminal cysteine, or of the entire carboxy-terminal region, may prevent or slow down correct formation of the inner shell, which may in turn prevent or slow down formation of the outer shell.

The COOH-terminus of the wild-type full-length protein is thought to be inside the

core particle (Zhou *et al.*, 1992) but the COOH-terminus of at least some of the fusion proteins is necessarily surface-accessible as the core fusion particles are bound by antibodies directed against the fusion section (Stahl and Murray, 1989). However, examination of the disulphide bond content of core fusion particles shows that the presence of the COOH-terminus on the surface of the particles also has an effect on disulphide bond formation. In addition, this effect is greater than that of the other mutants studied as the disulphide bond pattern is not that of wild-type, even after extended air oxidation.

The behaviour of HBcpreS1(1-20) once again indicates that the presence of a cysteine at the COOH-terminus is important and results in at least 80% disulphide bond formation between Cys48-Cys48 and/or Cys61-Cys61. It is expected that the cysteines linked by disulphide bonds in the fusion proteins would be the same as those linked in the wild-type protein as the fusion proteins are known to react with anti-HBcAg antibodies to the same extent as wild-type HBcAg (Stahl and Murray, 1989). HBcpreS1(1-36), a protein with a tail of similar length but lacking a COOH-terminal cysteine existed only as monomer even after extended air oxidation. This was despite the tail being non-arginine-rich like that of the wild-type HBcAg. However, HBcE46, which has a tail slightly shorter than that of the wild-type does form a small proportion (6%) of dimer, and hence Cys48-Cys48 and/or Cys61-Cys61 bonds. However, this was the percentage formed after extended air oxidation and so it is not directly analagous to the situation with the wild-type truncated proteins.

Thus, both the size of the COOH tail and its position in the particle are important with regard to disulphide bond formation in the remainder of the molecule. If dimers are not tightly linked by Cys183-Cys183 bonds but are more loosely associated, although still forming particles, this may prevent the remainder of the molecules from being in the correct position and/or conformation to form Cys48-Cys48 and/or Cys61-Cys61 bonds with neighbouring molecules to create the approximately spherical particle. This effect is reduced as the length of the COOH-terminal tail is reduced, perhaps because it then causes less of an obstruction.

A COOH tail displayed on the surface of the particle also appears to be more disruptive than one present internally. It may be that as the mutations in the non-fusion proteins are either subtle (C183S) or remove only a part of the protein known not to be essential for core particle formation, the overall effect of these mutations will be smaller than those in the fusion proteins. In the latter proteins the region known not to be important for particle formation is not merely removed but is substituted by a large number of foreign amino acids. This has been shown, previously, to alter the particle structure by presenting this region of the protein on the surface rather than the interior of the particle, in at least a proportion of the proteins (Stahl and Murray, 1989).

The work presented here now demonstrates that there are further repercussions in the HBcAg-derived remainder of the fusion proteins in terms of disulphide bond formation, which reflects a difference in structure and stability. However, these differences can be compensated if a cysteine is present at the COOH-terminus of the fusion protein in which case the fusion region is still displayed, at least in part, on the particle surface, but the structure of the remainder of the particle is not affected and the particles are stable. As stability of recombinant vaccines is desirable, the inclusion of a cysteine residue at the COOH-terminus of the protein would be predicted to allow disulphide bond formation to occur to the extent of the wild-type HBcAg throughout the entire particle. This is likely to be true of most COOH-terminal fusion sequences as the highly-charged arginine-rich terminus of the wild-type protein, and the unremarkable preS1(1-20) sequences reacted in the same manner, with regard to disulphide bond formation, upon inclusion of a carboxy-terminal cysteine. Similarly, N-terminal fusions (Clarke *et al.*, 1987; Beesley *et al.*, 1990; Francis *et al.*, 1990; Schoedel *et al.*, 1992) and fusions within HBcAg (Borisova *et al.*, 1989; Schoedel *et al.*, 1992) may be made more stable by the retention of the native COOH tail with its cysteine residue.

CHAPTER 6: Conclusions

Work presented in this thesis has demonstrated that the amino- and carboxy-terminal regions of HBcAg and HBcAg-derived proteins are important for the structural integrity of the particles they form and that the disulphide bond pattern of HBcAg, determined during this work, is not essential for formation of core particles.

The aim of the work presented in Chapter 3 of this thesis was to determine the reason for the low efficiency of interaction with antibody of the ammonium sulphate-precipitated extract of the product of pRI-4 relative to the protein product of pR1-11, when the proteins differed by only two amino acids (amino acids 3 and 4 of HBcAg). The reasons for this were found to be two-fold. Firstly, the amount of each protein present differed and this is an example of the sensitivity of HBcAg to perturbation of its coding sequence, when expressed in *E. coli,* which has also been noted by others. Interestingly, a gradation of the amount of protein present was found in the panel of amino-terminus mutant proteins examined, from normal levels of protein to no detectable protein, with reduction in the levels of protein present having occurred only when amino acids 3, 4 or 5 were mutated. While mRNA was detected in all cases, even when protein was not detected, there may be different levels of core-specific mRNA present in cells transfected with different mutants as the PCR-based assay used is highly sensitive but not quantitative. Alternatively, there may be reduced efficiency of translation of certain mutants, possibly caused by unusual RNA structures. Finally, the protein products may be undesirable to the cell and consequently degraded. In summary, it is clear that levels of HBcAg-derived proteins produced in *E. coli* can vary enormously due to only minor alterations to the coding sequence.

The role of the amino-terminus of HBcAg in core particle structure determination was also shown in Chapter 3. Proteins lacking amino acid 3 reacted less well with polyclonal anti-HBcAg/anti-HBeAg antibodies, relative to wild-type, while the other mutants examined reacted in a manner similar to wild-type. This may mean that deletion of amino acid 3 causes disruption of the structure of regions of HBcAg that react with antibody, either indirectly, or directly if amino acid 3 is contained within an antibody-binding site. In summary, the retention of amino acid 3 and possibly

amino acid 4, at least within the context of the proteins studied in this work, is not important for core particle formation but is important for the structure of these particles.

Further investigation of the structure of core particles was presented in Chapter 4. The elucidation of the disulphide bond pattern of HBcAg during this work was used as a means of gaining information about the core particle structure, as cysteine residues between which a disulphide bond forms are necessarily very close. The following conclusions were drawn:

1. 100% of Cys 61 and at least 50% of Cys48 are at the interface between HBcAg subunits of core particles, as they form disulphide bonds with the equivalent cysteine residue on an adjacent molecule.

2. HBcAg monomers appear to exist in at least two structural forms within the particle as approximately 50% of Cys48 is present in a disulphide bond while the remaining 50% is not involved in a disulphide bond. Therefore each Cys48 is not in the same position in each molecule.

3. Cys183 is also at an interface between HBcAg monomers but as it is thought to be inside the particle it may form an "inner shell" similar to that of the nucleocapsid of Tomato Bushy Stunt Virus. If this is indeed the case, it may also have the same function, in aiding correct formation of the "outer shell" of the nucleocapsid, as certain of the mutant proteins with an altered carboxy-terminus formed disulphide bonds in the remainder of the structure less readily than wild-type HBcAg.

4. Cys107 may or may not be at an interface between HBcAg monomers. However, it can be said that it is not in the correct position relative to any other cysteines of the monomer or particle, to be able to form a disulphide bond with them.

This work was taken further in the work presented in Chapter 5. Examination of the

disulphide bond pattern of core fusion proteins, once again using disulphide bond formation as an indicator of protein structure, showed that the nature of the carboxy-terminal region of HBcAg has an effect on the fine structure of the remainder of the particle. While the reaction of anti-HBcAg antibodies with these proteins is not affected, relative to wild-type HBcAg (Stahl and Murray, 1989), fine structural alterations are apparent as there is a difference in disulphide bond formation. Thus, while particles of apparently normal morphology are formed by all fusion proteins examined, the structure of the particle is in fact affected by the nature of the carboxy-terminus region. Possibly, in the fusion proteins in which the normal disulphide bond complement is not formed, the monomers have structure very similar to wild-type, and so react with antibody, but the monomers are held sufficiently apart by the abberrant carboxy-terminal region to prevent the close contact of the monomers in the core particle.

In conclusion, the elucidation of the disulphide bond pattern of wild-type HBcAg has allowed more general conclusions concerning the particle structure to be drawn by investigation of disulphide bond formation in mutant proteins. While the stability of the structure is enhanced by the cysteine residues of HBcAg, the determination of the structure does not require these residues. As mutation of a cysteine codon to a serine codon requires only one nucleotide change, the fact that this has not occurred *in vivo* means that the presence of cysteine residues is essential for the virus. This work has excluded some possible reasons for this. Those that remain to be investigated include a role in the function of HBeAg, a role in infection of cells, or a role in maintaining the free energy of the particle at the optimum value for virus survival, which could be investigated only by examination of the viability of a virus lacking cysteines after very many rounds of infection.

References

Acs, G., Sells, M.A., Purcell, R.H., Price, P., Engle, R., Shapiro, M., Popper, H. (1987). Hepatitis B virus produced by transfected HepG2 cells causes hepatitis in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 4641-4644.

Aden, D.P., Fogel, A., Plotkin, S., Damjanov, I., Knowles, B.B. (1979). Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature (London)* **282**: 615-616.

Albin, C. and Robinson, W.S. (1980). Protein kinase activity in hepatitis B virus. *Journal of Virology* **34**: 297-302.

Antonucci, T.K. and Rutter, W.J. (1989). Hepatitis B virus (HBV) promoters are regulated by the HBV enhancer in a tissue-specific manner. *Journal of Virology* **63**: 579-583.

Ashton-Rickardt, P.G. (1988). PhD Thesis. University of Edinburgh.

Aufiero, B. and Schneider, R.J. (1990). The hepatitis B virus X gene product *trans*-activates both RNA polymerase II and III promoters. *EMBO Journal* **9**: 497-504.

Bachmann, B.J. (1972). *Bacteriology Reviews* **36**: 525-557.

Barker, L.F., Almeida, J.D., Hoofnagle, J.H., Gerety, R.J., Jackson, D.R., McGrath, P.P. (1974). Hepatitis B core antigen: immunology and electron microscopy. *Journal of Virology* **14**: 1552-1558.

Barry, T., Geary, S., Hannify, S., MacGearailt, C., Shalloo, M., Heery, D., Gannon, F., Powell, R. (1992). Rapid mini-preparations of total RNA from bacteria. *Nucleic Acids Research* **20**: 4940.

Bartenschlager, R. and Schaller, H. (1988). The amino-terminal domain of the hepadnaviral P gene encodes the terminal protein (genome-linked protein) believed to prime reverse transcription. *EMBO Journal* **7**: 4185-4192.

Bartenschlager, R., Junker-Niepmann, M., Schaller, H. (1990). The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *Journal of Virology* **64**: 5324-5332.

Bartenschlager, R. and Schaller, H. (1992). Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO Journal* **11**: 3413-3420.

Bavand, M., Feitelson, M., Laub, O. (1989). The hepatitis B virus-associated reverse transcriptase is encoded by the viral polymerase gene. *Journal of Virology* **63**: 1019-1021.

Bayer, M., Blumberg, B., Werner, B. (1968). Particles associated with Australia antigen in the sera of patients with leukemia, Down's Syndrome and hepatitis. *Nature (London)* **218**: 1057-1059.

Bchini, R., Capel, F., Daugnet, C., Dubanchet, S., Petit, M.-A. (1990). *In vitro* infection of human hepatoma (HepG2) cells with hepatitis B virus. *Journal of Virology* **64**: 3025-3032.

Beasley, R.P., Hwang, L.-Y., Lin, C.-C., Chien, C.-S. (1981). Hepatocellular carcinoma and hepatitis B virus. *Lancet* (ii) 1129-1133.

Beesley, K.M., Francis, M.J., Clarke, B.E., Beesley, J.E., Doppinghepenstal, P.J.C., Clare, J.J., Brown, F., Romanos, M.A. (1990). Expression in yeast of amino-terminal peptide fusions to hepatitis B core antigen and their immunological properties. *Bio-Technology* **8**: 644-649.

Bhatnagar, P.K., Papas, E., Blum, H.E., Milich, D.R., Nitecki, D., Karels, M.J., Vyas, G.N. (1982). Immune response to synthetic peptide analogs of hepatitis B surface antigen specific for the alpha determinant. *Proceedings of the National Academy of Sciences of the United States of America* **79**: 4400-4404.

Birnbaum, F. and Nassal, M. (1990). Hepatitis B virus nucleocapsid assembly - primary structure requirements in the core protein. *Journal of Virology* **64**: 3319-3330.

Blum, H.E., Stowring, L., Figus, A., Montgomery, C.K., Haase, A.T., Vyas, G.N. (1983). Detection of hepatitis B virus DNA in hepatocytes, bile duct epithelium and vascular elements by *in situ* hybridisation. *Proceedings of the National Academy of Sciences of the United States of America* **80**: 6685-6688.

Blum, H.E., Zhang, Z.-S., Galun, E., vonVeizsacker, F., Garner, B., Liang, J.R. (1992). Hepatitis B virus X protein is not central to the viral life cycle *in vitro*. *Journal of Virology* **66**: 1223-1227.

Blumberg, B.S., Alter, H.J., Visnich, S.J. (1965), A "new" antigen in leukemia sera. *Journal of the American Medical Association* **191**: 541-546.

Borisova, G.P., Berzins, I., Pushko, P.M., Pumpen, P., Gren, E.J., Tsibinogin, V.V., Loseva, V., Ose, V., Ulrich, R., Siakkou, H., Rosenthal, H.A. (1989). Recombinant core particles of hepatitis B virus exposing foreign antigenic determinants on their surface. *FEBS Letters* **259**: 121-124.

Bosch, V., Bartenschlager, R., Radziwill, G., Schaller, H. (1988). The duck hepatitis virus P gene codes for protein strongly associated with the 5′ end of the viral DNA minus strand. *Virology* **166**: 475-486.

Brady, J., Radonovich, M., Vodkin, M., Natarajan, V., Thoren, M., Das, G., Janik, J., Salzman, N.P. (1982). Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA. *Cell* **31**: 625-633.

Brown, A.L., Francis, M.J., Hastings, G.Z., Parry, N.R., Barnett, P.V., Rowlands, D.J., Clarke, B.E. (1991). Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted. *Vaccine* **9**: 595-601.

Bruss, V. and Gerlich, W.H. (1988). Formation of transmembrane hepatitis B E antigen by cotranslational *in vitro* processing of the viral precore protein. *Virology* **163**: 268-275.

Bruss, V. and Ganem, D. (1991). The role of envelope proteins in hepatitis B virus assembly. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 1059-1063.

Budkowska, A., Kalinowska, B., Nowaslawski, A. (1979). Identification of two HBeAg subspecificities revealed by chemical treatment and enzymatic digestion of liver-derived HBcAg. *Journal of Immunology* **123**: 1415-1416.

Bulla, G.A. and Siddiqui, A. (1988). The hepatitis B virus enhancer modulates transcription of the hepatitis B virus surface antigen gene from an internal location. *Journal of Virology* **62**: 1437-1441.

Bulla, G.A. and Siddiqui, A. (1989). Negative regulation of the hepatitis B virus preS1 promoter by internal DNA sequences. *Virology* **170**: 251-260.

Burrell, C.J., MacKay, P., Greenaway, P.J., Hofschneider, P.H., Murray, K. (1979). Expression in *Escherichia coli* of hepatitis B virus DNA sequences cloned in plasmid pBR322. *Nature (London)* **279**: 43-47.

Caselmann, W.H., Meyer, M., Kekule, A.S., Lauer, U., Hofschneider, P.H., Koshy, R. (1990). A trans-activator function is generated by integration of hepatitis B virus preS/S sequences in human hepatocellular carcinoma DNA. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 2970-2974.

Cattaneo, R., Will, H., Hernandez, N., Schaller, H. (1983). Signals regulating hepatitis B surface antigen transcription. *Nature (London)* **305**: 336-338.

Cattaneo, R., Will, H., Schaller, H. (1984). Hepatitis B virus transcription in the infected liver. *EMBO Journal* **3**: 2191-2196.

Chakraborty, P.R., Ruiz-Opazo, N., Shouval, D., Shafritz, D.A. (1980). Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular cell line. *Nature (London)* **286**: 531-533.

Chang, C., Jeng, K., Hu, C., Lo, S.J., Su, T., Ting, L.P., Chou, C.K., Han, S., Pfaff, E., Salfeld, J., Schaller, H. (1987). Production of hepatitis B virus *in vitro* by transient expression of cloned HBV DNA in a hepatoma cell line. *EMBO Journal* **6**: 675-680.

Chang, H.-K., Wang, B.-Y., Yuh, C.-H., Wei, C.-L., Ting, L.-P. (1989a). A liver-specific nuclear factor interacts with the promoter region of the large surface protein gene of human hepatitis B virus. *Molecular and Cellular Biology* **9**: 5189-5197.

Chang, L.-J., Pryciak, P., Ganem, D. (1989b). Biosynthesis of the reverse transcriptase of hepatitis B viruses involves *de novo* translational initiation not ribosomal frameshifting. *Nature (London)* **337**: 364-368.

Chang, L.-J., Hirsch, R.C., Ganem, D., Varmus, H.E. (1990). Effects of insertional and point mutations on the functions of the duck hepatitis B virus polymerase. *Journal of Virology* **64**: 5553-5558.

Chemello, L., Pontisso, P., Schiavon, E., Thiers, V., Tagariello, G., Alberti, A. (1988). Hepatitis B core antigen in serum during acute hepatitis B. *Journal of Medical Virology* **24**: 361-367.

Chen, P.J., Chen, C.R., Sung, J.L., Chen, D.S. (1989). Identification of a doubly-spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. *Journal of Virology* **63**: 4165-4171.

Chen, Y., Robinson, W.S., Marion, P.L. (1992). Naturally-occurring point mutation in the C-terminus of the polymerase gene prevents duck hepatitis B virus RNA packaging. *Journal of Virology* **66**: 1282-1287.

Cherrington, J., Russnak, R., Ganem, D. (1992). Upstream sequences and cap proximity in the regulation of polyadenylation in ground squirrel hepatitis virus. *Journal of Virology* **66**: 7589-7596.

Chiang, P.-W., Hu, C.-P., Su, T.-S., Lo, S.-J., Chu, M.-H.H., Schaller, H., Chang, C.-M. (1990). Encapsidation of truncated human hepatitis B virus genomes through trans-complementation of the core protein and polymerase. *Virology* **176**: 355-361.

Chiang, P.-W., Jeng, K.-S., Hu, C.-P., Chang, C.-M. (1992). Characterisation of a *cis* element required for packaging and replication of the human hepatitis B virus. *Virology* **186**:701-711.

Chisaka, O., Araki, K., Ochiya, T., Tsurimoto, T., Hiranyawasitte-Attatippaholkun, W., Yanaihara, N., Matsubara, K. (1987). Purification of hepatitis B virus X product synthesised in *Escherichia coli* and its detection in a human hepatoblastoma cell line producing hepatitis B virus. *Gene* **60**: 183-189.

Churchhill, M.E.A. and Travers, A.A. (1991). Protein motifs that recognise structural features of DNA. *Trends in Biochemical Sciences* **16**: 92-97.

Clarke, B.E., Newton, S.E., Carroll, A.R., Francis, M.J., Appleyard, G., Syred, A.D., Highfield, P.E., Rowlands, D.J., Brown, F. (1987). Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. *Nature (London)* **330**: 381-384.

Clarke, B.E., Brown, A.L., Grace, K.G., Hastings, G.Z., Brown, F., Rowlands, D.J., Francis, M.J. (1990). Presentation and immunogenicity of viral epitopes on the surface of hybrid hepatitis B virus core particles produced in bacteria. *Journal of General Virology* **71**: 1109-1117.

Cohen, B.J. (1978). *Journal of Medical Virology* **3**: 141.

Cohen, B.J. and Richmond, J.E. (1982). Electron microscopy of a hepatitis B core antigen synthesised in *E. coli. Nature (London)* **296**: 677-678.

Colgrove, R., Simon, G., Ganem, D. (1989). Transcriptional activation of homologous and heterologous genes by the hepatitis B virus X gene product in cells permissive for viral replication. *Journal of Virology* **63**: 4019-4026.

Creighton, T.E. (1984a). Disulphide bond formation in proteins. *Methods in Enzymology* **107**: 305-329.

Creighton, T.E. (1984b). Disulphide bonds and protein stability. *BioEssays* **8**: 57-63.

Dane, D.S., Cameron, C.H., Briggs, M. (1970). Virus-like particles in serum hepatitis patients with Australia antigen-associated hepatitis. *Lancet* **1970(i)**: 695-698.

DeLoia, J.A., Burk, R.D., Gearhart, J.D. (1989). Developmental regulation of hepatitis B surface antigen expression in two lines of hepatitis B virus transgenic mice. *Journal of Virology* **63**: 4069-4073.

DeMedina, T., Faktor, O., Shaul, Y. (1988). The S promoter of hepatitis B virus is regulated by positive and negative elements. *Molecular and Cellular Biology* **8**: 2449-2455.

Dudley, F., Fox, R., Sherlock, S. (1972). *Lancet* (i): 763-766.

Eckhardt, S.G., Milich, D.R., McLachlan, A. (1991). Hepatitis B virus core antigen has 2 nuclear localisation sequences in the arginine-rich carboxyl terminus. *Journal of Virology* **65**: 575-582.

Edman, J.C., Gray, P., Valenzuela, P., Rall, L.B., Rutter, W.J. (1980). Integration of hepatitis B virus sequences and their expression in a human hepatoma cell. *Nature (London)* **286**: 535-538.

Edman, J.C., Hallewell, R.A., Valenzuela, P., Goodman, H.M., Rutter, W.J. (1981). Synthesis of hepatitis B surface and core antigens in *E.coli. Nature* **291**: 503-506.

Emini, E.A., Hughs, J.V., Perlow, D.S., Boger, J. (1989). Induction of hepatitis B virus-neutralising antibody by a virus-specific synthetic peptide. *Journal of Medical Virology* **28**: 7-12.

Enders, G.H., Ganem, D., Varmus, H.E. (1985). Mapping the major transcripts of ground squirrel hepatitis virus: the presumptive template for reverse transcriptase is terminally redundant. *Cell* **42**: 297-308.

Enders, G.H., Ganem, D., Varmus, H.E. (1987). 5'-terminal sequences influence the segregation of ground squirrel hepatitis virus RNAs into polyribosomes and viral core particles. *Journal of Virology* **61**: 35-41.

Faktor, O. and Shaul, Y. (1990). The identification of hepatitis B virus X gene responsive elements reveals functional similarity of X and HTLV-1 *tax*. *Oncogene* **5**: 867-872.

Feitelson, M.A., Marion, P.L., Robinson, W.S. (1982). Core particles of hepatitis B virus and ground squirrel hepatitis virus II: characterization of the protein kinase reaction associated with ground squirrel hepatitis virus and hepatitis B virus. *Journal of Virology* **43**: 741-748.

Feitelson, M.A. and Miller, R.H. (1988). X gene-related sequences in the core gene of duck and heron hepatitis B viruses. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6162-6166.

Feitelson, M.A. (1989). Hepatitis B virus gene products as immunological targets in chronic infection. *Mol Biol Med* **6**: 367-393.

Feitelson, M.A., Clayton, M.M., Phimister, B. (1990). Monoclonal antibodies raised to purified woodchuck hepatitis virus core antigen particles demonstrate X antigen reactivity. *Virology* **177**: 357-366.

Fernholz, O., Stemler, M., Brunetto, M.R., Bonino, F., Will, H. (1991). *Journal of Hepatology* **13**: 102-104.

Francis, M.J. and Clarke, B.E. (1989). Peptide vaccines based on enhanced immunogenicity of peptide epitopes presented with T-cell determinants of hepatitis B core protein. *Methods in Enzymology* **178**: 659-676.

Francis, M.J., Hastings, G.Z., Brown, A.L., Grace, K.G., Rowlands, D.J., Brown, F., Clarke, B.E. (1990). Immunological properties of hepatitis B core antigen fusion proteins. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 2545-2549.

Freedman, R.B. (1984). Native disulphide bond formation in protein biosynthesis: evidence for the role of protein disulphide isomerase. *Trends in Biochemical Sciences* **9**: 438-441.

Fujiyama, A., Miyanohara, A., Nozaki, C., Yoneyama, T., Ohtomo, N., Matsubara, K. (1983). Cloning and structural analyses of hepatitis B virus DNAs, subtype *adr*. *Nucleic Acids Research* **11**: 4601-4610.

Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P., Charnay, P. (1979). Nucleotide sequence of the hepatitis B virus genome (subtype *ayw*) cloned in *E. coli*. *Nature (London)* **281**: 646-650.

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., Milanesi, G. (1989). A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. *Journal of Virology* **63**: 4645-4652.

Galun, E., Offensperger, W.B., Vonweizsacker, F., Offensperger, S., Wands, J.R., Blum, H.E. (1992). Human non-hepatocytes support hepadnaviral replication and virion production. *Journal of General Virology* **73**: 173-178.

Ganem, D. (1982). Persistent infection of humans with hepatitis B virus: mechanisms and consequences. *Reviews of Infectious Diseases* **4**: 1026-1047.

Ganem, D. and Varmus, H.E. (1987). The molecular biology of the hepatitis B viruses. *Annual Review of Biochemistry* **56**: 651-693.

Ganem, D. (1991). Assembly of hepadnaviral virions and subviral particles. *Current Topics in Microbiology and Immunology* **168**: 61-84.

Garcia, P., Ou, J., Rutter, W.J., Walter, P. (1988). Targetting of protein of hepatitis B virus to the endoplasmic reticulum membrane: after signal peptide cleavage translocation can be aborted and the product released into the cytoplasm. *Journal of Cell Biology* **106**: 1093-1104.

Gerin, J., Purcell, R., Hoggan, M., Holland, P., Chanock, R. (1969). Biophysical properties of Australia antigen. *Journal of Virology* **4**: 763-768.

Gerin, J.L., Alexander, H., Shih, J.W-.K., Purcell, R.H., Dopolito, G., Engle, R., Green, N., Sutcliffe, J.G., Shinnick, T.M., Lerner, R.A. (1983). Chemically synthesised peptides of HBsAg duplicate the *d/y* subtype specificities and induce subtype specific antibodies in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America* **80**: 2365-2369.

Gerlich, W.H. and Robinson, W.S. (1980). Hepatitis B virus contains protein attached to the 5′ terminus of its complete DNA strand. *Cell* **21**: 801-809.

Gerlich, W., Goldmann, U., Muller, R., Stibbe, W., Wolff, W. (1982). Specificity and localization of the hepatitis B virus-associated protein kinase. *Journal of Virology* **42**: 761-766.

Gocke, D. (1975). Extrahepatic manifestation of viral hepatitis. *American Journal of Medical Science* **207**: 49-55.

Gough, N. and Murray, K. (1982). Expression of the hepatitis B virus surface, core and E antigen genes by stable rat and mouse cell lines. *Journal of Molecular Biology* **162**: 43-67.

Gough, N.M. (1983). Core and E antigen synthesis in rodent cells transformed with hepatitis B virus DNA is associated with greater than genome length viral messenger RNAs. *Journal of Molecular Biology* **165**: 683-699.

Guo, W.T., Chen, M., Yen, T.S.B, Ou, J.H. (1993). Hepatocyte-specific expression of the hepatitis B virus core promoter depends on both positive and negative regulation. *Molecular and Cellular Biology* **13**: 443-448.

Harrison, S. (1985). Principles of virus structure. In *Virology*, Fields, B.N. ed. (Raven Press, New York) pp. 27-44.

Hatton, T., Zhou, S.L., Standring, D.N. (1992). RNA-binding and DNA-binding activities in hepatitis B virus capsid protein - a model for their roles in viral replication. *Journal of Virology* **66**: 5232-5241.

Heermann, K.H., Goldmann, U., Schwartz, W., Seyffarth, T., Baumgarten, H., Gerlich, W.H. (1984). Large surface proteins of hepatitis B virus containing the preS sequence. *Journal of Virology* **52**: 396-402.

Heermann, K.H., Kruse, F., Seifer, M., Gerlich, W.H. (1987). Immunogenicity of the gene S and preS domains in hepatitis B virions and HBsAg filaments. *Intervirology* **28**: 14-25.

Hilditch, C.M., Rogers, L.J., Bishop, D.H.L. (1990). Physicochemical analysis of the hepatitis B virus core antigen produced by a baculovirus expression vector. *Journal of General Virology* **71**: 2755-2759.

Hilleman, M.R., Bertland, A.U., Buynak, E.B., Lampson, G.P., McAleer, W.J., McLean, A.A., Roehm, R.R., Tytell, A.A. (1978). Clinical and laboratory studies of HBsAg vaccine. In *Viral Hepatitis*, Vyas GN, Cohen SN, Schmid R eds. (Franklin Institute Press, Philadelphia) pp. 525-537.

Hirsch, R., Lavine, J., Chang, L., Varmus, H., Ganem, D. (1990). Polymerase gene products of hepatitis B viruses are required for genomic RNA packaging as well as for reverse transcription. *Nature (London)* **344**: 522-525.

Hirschman, R.J., Schulman, N.R., Barker, L.F., Smith, K.O. (1969). *Journal of the American Medical Association* **208**: 1667-1670.

Honigwachs, J., Faktor, O., Dickstein, R., Shaul, Y., Laub, O. (1989). Liver-specific expression of hepatitis B virus is determined by the combined action of the core gene promoter and the enhancer. *Journal of Virology* **63**: 919-924.

Hoofnagle, J.H., Gerety, R.J., Barker, L.F, (1973). Antibody to hepatitis B virus core in man. *Lancet* (ii): 869-873.

Howe, A.Y.M., Elliott, J.F., Tyrrell, D.L.J. (1992). Duck hepatitis B virus polymerase produced by *in vitro* transcription and translation possesses DNA polymerase and reverse transcriptase activities. *Biochemical and Biophysical Research Communications* **189**: 1170-1176.

Hruska, J.F. and Robinson, W.S. (1977). The proteins of hepatitis B Dane particle cores. *Journal of Medical Virology* **1**: 119-131.

Hu, K.Q. and Siddiqui, A. (1991). Regulation of the hepatitis B virus gene expression by the enhancer element I. *Virology* **181**: 721-726.

Huan, B.F. and Siddiqui, A. (1992). Retinoid X receptor RXR-alpha binds to and transactivates the hepatitis B virus enhancer. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 9059-9063.

Imai, M., Yanase, Y., Nojiri, T., Miyakawa, U., Mayumi, M. (1979). A receptor for polymerised human and chimpanzee albumins on hepatitis B virus particles co-occuring with HBeAg. *Gastroenterology* **76**: 242-247.

Imamura, T., Suguhara, K., Adachi, S., Miyatsu, Y., Mizokami, H., Matsusaka, T. (1988). Purification and characterisation of the hepatitis B virus core antigen produced in the yeast *Saccharomyces cerevisiae*. *Journal of Biotechnology* **8**: 149-161.

Iwarson, S., Tabor, E., Thomas, H.C., Goodall, A., Waters, J., Snoy, P., Shih, J.W., Gerety, R.J. (1985a). Neutralization of hepatitis B virus infectivity by a murine monoclonal antibody: an experimental study in the chimpanzee. *Journal of Medical Virology* **16**: 89-96.

Iwarson, S., Tabor, E., Thomas, H.C., Snoy, P., Gerety, R.J. (1985b). Protection against hepatitis B virus infection by immunization with hepatitis B core antigen. *Gastroenterology* **88**: 763-767.

Jean-Jean, O., Levrero, M., Will, H., Perricaudet, M., Rossignol, J.M. (1989a). Expression mechanisms of the hepatitis B virus (HBV) C gene and biosynthesis of HBe antigen. *Virology* **170**: 99-106.

Jean-Jean, O., Salhi, S., Carlier, D., Elie, C., deRecondo, A.M., Rossignol, J.M. (1989b). Biosynthesis of hepatitis B virus E antigen - directed mutagenesis of the putative aspartyl protease site. *Journal of Virology* **63**: 5497-5500.

Jean-Jean, O., Wiemer, T., deRecondo, A.M., Will, H., Rossignol, J.M. (1989c). Internal entry of ribosomes and ribosomal scanning involved in hepatitis B virus P gene expression. *Journal of Virology* **63**: 5451-5454.

Jeng, K.S., Hu, C.P., Chang, C.M. (1991). Differential formation of disulfide linkages in the core antigen of extracellular and intracellular hepatitis B virus core particles. *Journal of Virology* **65**: 3924-3927.

Junker-Niepmann, M., Bartenschlager, R., Schaller, H. (1990). A short *cis*-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO Journal* **9**: 3389-3396.

Kaneko, S. and Miller, R.H. (1988). X region-specific transcript in mammalian hepatitis B-infected liver. *Journal of Virology* **62**: 3979-3984.

Kaplan, P.M., Greenman, R.L., Gerin, J.L., Purcell, R.H., Robinson, W.S. (1973). DNA polymerase associated with human hepatitis B antigen. *Journal of Virology* **12**: 995-1005.

Karpen, S., Banerjee, R., Zelent, A., Price, P., Acs, G. (1988). Identification of protein-binding sites in the hepatitis B virus enhancer and core promoter domains. *Molecular and Cellular Biology* **8**: 5159-5165.

Kawamoto, S., Yamamoto, S., Ueda, K., Nagahata, T., Chisaka, O., Matsubara, K. (1990). *Biochemical and Biophysical Research Communications* **171**: 1130-1136.

Kay, A., Mandart, E., Trepo, C., Galibert, F. (1985). The HBV HBx gene expressed in *E. coli* is recognised by sera from hepatitis patients. *EMBO Journal* **4**: 1287-1292.

Kekule, A.S., Lauer, U., Meyer, M., Caselmann, W.H., Hofschneider, P.H., Koshy, R. (1990). The preS2/S region of integrated hepatitis B virus DNA encodes a transcriptional transactivator. *Nature (London)* **343**: 457-461.

Kekule, A.S., Lauer, U., Weiss, L., Luber, B., Hofschneider, P.H. (1993). Hepatitis B virus transactivator HBx uses a tumor promoter signalling pathway. *Nature (London)* **362**: 742-745.

Khudyakov, Y.E. and Makhov, A.M. (1989). Prediction of terminal protein and ribonuclease H domains in the gene P product of hepadnaviruses. *FEBS Letters* **243**: 115-118.

Kniskern, P.J., Hagopian, A., Montgomery, D.L., Burke, P., Dunn, N.R., Hofmann, K.J., Miller, W.J., Ellis, R.W. (1986). Unusually high expression of a foreign gene (hepatitis B virus core antigen) in *Saccharomyces cerevisiae. Gene* **46**: 135-141.

Kobayashi, M. and Koike, K. (1984). Complete nucleotide sequence of hepatitis B virus DNA of subtype *adr* and its conserved gene organization. *Gene* **30**: 227-232.

Koike, K., Shirakata, Y., Yaginuma, K., Arii, M., Takada, S., Nakamura, I., Hayashi, Y., Kawada, M., Kobayashi, M. (1989). Oncogenic potential of hepatitis B virus. *Molecular Biology and Medicine* **6**: 151-160.

Kojima, T., Bloemen, J., Desmet, V.J. (1987). Immune electron microscopic demonstration of hepatitis B core antigen (HBcAg) in liver cell plasma membranes. *Liver* **7**: 191-200.

Korba, B.E., Cote, P.J., Gerin, J.L. (1988). Mitogen-induced replication of woodchuck hepatitis virus in cultured peripheral blood lymphocytes. *Science* **241**: 1213-1216.

Korba, B.E. et al (1989). Natural history of woodchuck hepatitis virus infections during the course of experimental infection: molecular virologic features of the liver and lymphoid tissues. *Journal of Virology* **63**: 1360-1370.

Koshy, R. and Hofschneider, P.H. (1989). Transactivation by hepatitis B virus may contribute to hepatocarcinogenesis. In *Current Topics in Microbiology and Immunology* **144**: 265-281.

Kramer, B., Kramer, W., Fritz, H.J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell* **38**: 879-887.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**: 680-685.

Landers, T.A., Greenberg, H.B., Robinson, W.S. (1977). The structure of hepatitis B virus DNA and nature of the endogenous DNA polymerase reaction. *Journal of Virology* **23**: 368-376.

Lanzavecchia, A. (1985). Antigen-specific interaction between T and B cells. *Nature (London)* **314**: 537-539.

Laub, O., Rall, L.B., Truett, M., Shaul, Y., Standring, D.N., Valenzuela, P., Rutter (1983). Synthesis of hepatitis B surface antigen in mammalian cells: expression of the entire gene and the coding region. *Journal of Virology* **48**: 271-280.

LeBouvier, G.L. (1973). *Ann Intern Med* **79**: 894-896.

Lien, J.M., Aldrich, C.E., Mason, W.S. (1986). Evidence that a capped oligoribonucleotide is the primer for duck hepatitis B virus plus strand synthesis. *Journal of Virology* **57**: 229-237.

Liu, C.C, Yansura, D., Levinson, A. (1982). Direct expression of hepatitis B surface antigen in monkey cells from an SV40 vector. *DNA* **1**: 213-221.

Lucito, R. and Schneider, R.J. (1992). Hepatitis B virus X protein activates transcription factor NF KB without a requirement for protein kinase C. *Journal of Virology* **66**: 983-991.

McGlynn, E. and Murray, K. (1988). The hepatitis B virus polymerase: expression of its gene in *Escherichia coli*, and detection of antibodies to the product in convalescent sera. In *Viral Hepatitis and Liver Disease*, Zuckermann AJ ed. (Alan R. Liss, New York) pp. 323-329.

McGlynn, E., Reutener, S., Matter, A., Wildner, G., Will, H., Lydon, N.B. (1992). Hepatitis B virus polymerase gene - expression of the long open reading frame using the baculovirus expression system. *Journal of General Virology* **73**: 1515-1519.

MacKay, P., Lees, J., Murray, K. (1981). The conversion of hepatitis B core antigen synthesised in *E. coli* into E antigen. *Journal of Medical Virology* **8**: 237-243.

McLachlan, A., Milich, D.R., Raney, A.K., Riggs, M.G., Hughes, J.L., Sorge, J., Chisari, F.V. (1987). Expression of hepatitis B virus surface and core antigens: influences of preS and precore sequences. *Journal of Virology* **61**: 683-692.

Machida, A., Kishimoto, S., Ohumura, H., Miyamoto, H., Baba, K., Oda, K., Nakamura, T., Miyakawa, Y. (1983). A hepatitis B surface antigen polypeptide (P31) with the receptor for polymerised human as well as chimpanzee albumins. *Gastroenterology* **85**: 268-274.

Machida, A., Kishimoto, S., Ohnuma, H., Baba, K., Ito, Y., Miyamoto, H., Funatsu, G., Oda, K., Usuda, S., Togami, S., Nakamura, T., Miyakawa, Y., Mayumi, M. (1984). A polypeptide containing 55 amino acid residues coded by the preS region of hepatitis B virus deoxyribonucleic acid bears the receptor for polymerised human as well as chimpanzee albumins. *Gastroenterology* **86**: 910-918.

Machida, A., Ohnuma, H., Tsuda, F., Yoshikawa, A., Hoshi, Y., Tanaka, T., Kishimoto, S., Akahane, Y., Miyakawa, Y., Mayumi, M. (1991). Phosphorylation in the carboxyl terminal domain of the capsid protein of hepatitis B virus: Evaluation with a monoclonal antibody. *Journal of Virology* **65**: 6024-6030.

Magnius, L.O. and Espmark, J.A. (1972). New specificities in Australia antigen positive sera distinct from the LeBouvier determinants. *Jounral of Immunology* **109**: 1017-1021.

Marion, P.L., Oshiro, L.S., Regnery, D.C., Scullard, G.H., Robinson, W.S. (1980). A virus in Beechy ground squirrels which is related to hepatitis B virus of humans. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 2941-2945.

Mason, W.S., Seal, G., Summers, J. (1980). Virus of Pekin ducks with structural and biological relatedness to human hepatitis B virus. *Journal of Virology* **36**: 829-836.

Mason, W.S., Aldrich, C., Summers, J., Taylor, J.M. (1982). Asymmetric replication of duck hepatitis B virus DNA in liver cells: Free minus-strand DNA. *Proceedings of the National Academy of Sciences of the United States of America* **79**: 3997-4001.

Mason, W.S., Halpern, M.S., England, J.M., Seal, G., Egan, J., Coates, L., Aldrich, C., Summers, J. (1983). Experimental transmission of duck hepatitis B virus. *Virology* **131**: 375-384.

Mason, W.S., Taylor, J.M., Hull, R. (1987). Retroid virus genome replication. *Advances in Virus Research* **32**: 35-96.

Matsuda, K., Satoh, S., Ohori, H. (1988). DNA-binding activity of hepatitis B e antigen polypeptide lacking the protaminelike sequence of nucleocapsid protein of human hepatitis B virus. *Journal of Virology* **62**: 3517-3521.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**: 269-276.

Meyers, M.L., Trepo, L.V., Nath, N., Sninsky, J.J. (1986). Hepatitis B polypeptide X: expression in *Escherichia coli* and identification of specific antibodies in sera from hepatitis B virus-infected humans. *Journal of Virology* **57**: 101-109.

Midgely, C.A. and Murray, N.E. (1985). T4 polynucleotide kinase: cloning of the gene (*pse*T) and amplification of its product. *EMBO Journal* **4**: 2695-2703.

Milich, D.R., Thornton, G.B., Neurath, A.R., Kent, S.B., Michel, M.L., Tiollais, P., Chisari, F.V. (1985). Enhanced immunogenicity of the preS region of hepatitis B surface antigen. *Science* **228**: 1195-1198.

Milich, D.R. and McLachlan, A. (1986). The nucleocapsid of hepatitis B virus is both a T cell-independent and a T cell-dependent antigen. *Science* **234**: 1398-1401.

Milich, D.R. and McLachlan, A. (1987). A single 10-residue preS1 peptide can prime T cell help for antibody production to multiple epitopes within the preS1, preS2 and S regions of HBsAg. *Journal of Immunology* **138**: 4457-4465.

Milich, D.R., McLachlan, A., Moriarty, A., Thornton, G.B. (1987a). Immune response to hepatitis B virus core antigen (HBcAg) - localization of T-cell recognition sites within HBcAg/HBeAg. *Journal of Immunology* **139**: 1223-1231.

Milich, D.R., McLachlan, A., Thornton, G.B., Hughes, J.L. (1987b). Antibody production to the nucleocapsid and envelope of the hepatitis B virus primed by a single synthetic T-cell site. *Nature (London)* **329**: 547-549.

Milich, D.R., McLachlan, A., Stahl, S., Wingfield, P., Thornton, G.B., Hughes, J.L., Jones, J.E. (1988). Comparative immunogenicity of hepatitis B virus core and E antigens. *Journal of Immunology* **141**: 3617-3624.

Miller, R.H. (1987). Proteolytic self-cleavage of hepatitis B virus core protein may generate serum E antigen. *Science* **236**: 722-725.

Mishiro, S., Imai, M., Takahashi, K., Machida, A., Gotanda, T., Miyakawa, Y., Mayuni, M. (1980). A 49000-dalton polypeptide bearing all antigenic determinants and full immunogenicity of 22nm hepatitis B surface antigen particles. *Journal of Immunology* **124**: 1589-1593.

Miyaki, M., Sato, C., Gotanda, T., Matsui, T., Mishiro, S., Imai, M., Mayumi, M. (1986). Integration of region X of hepatitis B virus genome in human primary hepatocellular carcinomas propagated in nude mice. *Journal of General Virology* **67**: 1449-1454.

Miyanohara, A., Imamura, T., Araki, M., Sugawara, K., Ohtomo, N., Matsubara, K. (1986). Expression of hepatitis B virus core antigen gene in *Saccharomyces cerevisiae* - synthesis of 2 polypeptides translated from different initiation codons. *Journal of Virology* **59**: 176-180.

Molnar-Kimber, K.L., Summers, J., Taylor, J.M., Mason, W.S. (1983). Protein covalently bound to minus-strand DNA intermediates of duck hepatitis B virus. *Journal of Virology* **45**: 165-172.

Molnar-Kimber, K.L., Jarocki-Witck, V., Dheer, S.K., Vernon, S.K., Conley, A.J., Davis, A.R., Hung, P.P. (1988). Distinctive properties of hepatitis B virus envelope proteins. *Journal of Virology* **62**: 407-416.

Mondelli, M., Mieli, G.M., Alberti, A., Vergani, D., Portmann, B., Eddleston, A.L.W.F., Williams, R. (1982). Specificity of T lymphocyte cytotoxicity to autologous hepatocytes in chronic hepatitis B virus infection: evidence that T cells are directed against HBV core antigen expressed on hepatocytes. *Journal of Immunology* **129**: 2773-2778.

Moriarty, A.M., Alexander, H., Lerner, R.A. (1985). Antibodies to peptides detect new hepatitis B antigen: serological correlation with hepatocellular carcinoma. *Science* **227**: 429-433.

Murray, K., Bruce, S.A., Hinnen, A., Wingfield, P., vanErd, P.M.C.A., deReus, A., Schellekens, H. (1984). Hepatitis B virus antigens made in microbial cells immunise against viral infection. *EMBO Journal* **3**: 645-650.

Murray, K., Bruce, S.A., Wingfield, P., vanErd, P., deReus, A., Schellekens, H. (1987). Protective immunisation against hepatitis B with an internal antigen of the virus. *Journal of Medical Virology* **23**: 101-107.

Nakabayashi, H., Taketa, K., Miyano, K., Yamane, T., Sato, J. (1982). Growth of human hepatoma cell lines with differentiated functions in chemically defined medium. *Cancer Research* **42**: 3858-3863.

Nassal, M., Galle, P.R., Schaller, H. (1989). Protease-like sequence in hepatitis B virus core antigen is not required for E antigen generation and may not be part of an aspartic acid-type protease. *Journal of Virology* **63**: 2598-2604.

Nassal, M., Junker-Niepmann, M., Schaller, H. (1990). Translational inactivation of RNA function - discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell* **63**: 1357-1363.

Nassal, M. (1992a). The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *Journal of Virology* **66**: 4107-4116.

Nassal, M. (1992b). Conserved cysteines of the hepatitis B virus core protein are not required for assembly of replication-competent core particles nor for their envelopment. *Virology* **190**: 499-505.

Nassal, M., Rieger, A., Steinau, O. (1992). Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking. *Journal of Molecular Biology* **225**: 1013-1025.

Neurath, A.R. and Kent, S.B.H. (1985). In *Immunochemistry of Viruses*, eds. MHV van Regenmantel and AR Neurath, (Elsevier) pp325-366.

Neurath, A.R., Kent, S.B.H., Strick, N., Taylor, P., Stevens, C.E. (1985). Presence of preS gene-coded domains in hepatitis B virus (HBV) and their function. *Nature (London)* **351**: 154-156.

Neurath, A.R., Kent, S.B.H., Strick, N., Parker, K. (1986a). Identification and chemical synthesis of a host cell receptor binding site on a hepatitis B virus. *Cell* **46**: 429-436.

Neurath, A.R., Kent, S.B.H., Parker, K., Prince, A.M., Strick, N., Brotman, B., Sproul, P. (1986b). Antibodies to a synthetic peptide from preS 120-145 region of the hepatitis B virus envelope are virus-neutralising. *Vaccine* **4**: 35-37.

Neurath, A.R., Kent, S.B.H., Strick, N., Parker, K., Cuorouce, A.M., Riottot, M.M., Petit, M.A., Budkowska, A., Girard, M., Pillot, J. (1987). Antibodies to synthetic peptides from the preS1 and preS2 regions of one subtype of the hepatitis B virus (HBV) envelope protein recognize all HBV subtypes. *Molecular Immunology* **24**: 975-980.

Norrander, J., Kempe, T., Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**: 101-106.

Ogston, C.W., Schechter, E.M., Humes, C.A., Panikoff, M.B. (1989). Extrahepatic replication of woodchuck hepatitis virus in chronic infection. *Virology* **169**: 9-14.

Ohori, H., Onodera, S., Ishida, N. (1979). Demonstration of hepatitis B E antigen in association with intact Dane particles. *Journal of General Virology* **43**: 423-427.

Ohori, H., Yamaki, M., Onodera, S., Yamada, E., Ishida, N. (1980). Antigenic conversion of HBcAg to HBeAg by degradation of hepatitis B core particles. *Intervirology* **13**: 74-82.

Ohori, H. and Matsuda, K. (1989). Intracellular and extracellular distribution and immunochemical characterization of hepatitis B virus nucleocapsid proteins produced by a human hepatoma cell line transfected with cloned viral DNA. *Virology* **168**: 40-47.

Okada, K., Kamiyama, I., Inomata, I.M., Imai, M., Miyakawa, Y., Mayumi, M. (1976). E-antigen and anti-E antigen in the serum of asymptomatic carrier mothers as indicators of positive and negative transmission of hepatitis B virus to their infants. *New England Journal of Medicine* **294**: 746-749.

Okamoto, H., Imai, M., Shimozaki, M., Hoshi, Y., Iizuka, H., Gotanda, T., Tsuda, F., Miyakawa, Y., Mayumi, M. (1986). Nucleotide sequence of a cloned hepatitis B virus genome, subtype *ayr*: Comparison with genomes of the other three subtypes. *Journal of General Virology* **67**: 2305-2314.

Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugino, Y., Nishioka, K. (1983). Complete nucleotide sequences of cloned HNVB DNA; subtype *adr* and *adw*. *Nucleic Acids Research* **11**: 1747-1759.

Onodera, S., Ohori, H., Yamaki, M., Ishida, N. (1982). Electron microscopy of human hepatitis B virus cores by negative staining carbon film technique. *Journal of Medical Virology* **10**: 147-155.

Ou, J.H. and Rutter, W.J. (1985). Hybrid hepatitis B virus-host transcripts in a human hepatoma cell. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 83-87.

Ou, J.H., Laub, O., Rutter, W.J. (1986). Hepatitis B virus gene function - the precore region targets the core antigen to cellular membranes and causes the secretion of the E antigen. *Proceedings of the National Academy of Sciences of the United States of America* **83**: 1578-1582.

Ou, J.H. and Rutter, W.J. (1987). Regulation of secretion of the hepatitis B virus major surface antigen by the preS1 protein. *Journal of Virology* **61**: 782-786.

Ou, J.H., Yeh, C.T., Yen, T.S.B. (1989). Transport of hepatitis B virus precore protein into the nucleus after cleavage of its signal peptide. *Journal of Virology* **63**: 5238-5243.

Ou, J.H., Bao, H., Shih, C., Tahara, S.M. (1990). Preferred translation of human hepatitis B virus polymerase from core protein-specific but not from pre-core protein-specific transcript. *Journal of Virology* **64**: 4578-4581.

Overby, L.R., Hung, P.P., Lao, J.C.-H., Ling, C.M., Kekfuda, T. (1975). Rolling circular DNA associated with Dane particles in hepatitis B virus. *Nature (London)* **255**: 84-85.

Pasek, M., Goto, T., Gilbert, W., Zink, B., Schaller, H., MacKay, P., Leadbetter, G., Murray, K. (1979). Hepatitis B virus genes and their expression in *E. coli*. *Nature (London)* **282**: 575-579.

Peeples, M.E., Komai, K., Radek, R., Bankowski, M.J. (1987). A cultured cell receptor for the small S protein of hepatitis B virus. *Virology* **1**: 135-142.

Perfumo, S., Amicone, L., Colloca, S., Giorgio, M., Pozzi, L., Tripodi, M. (1992). Recognition efficiency of the hepatitis B virus polyadenylation signals is tissue-specific in transgenic mice. *Journal of Virology* **66**: 6819-6823.

Persing, D.H., Varmus, H.E., Ganem, D. (1985). A frameshift mutation in the preS region of the human hepatitis B virus genome allows production of surface antigen particles but eliminates binding to polymerized albumin. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 3440-3444.

Persing, D.H., Varmus, H.E., Ganem, D. (1987). The preS1 protein of hepatitis B virus is acylated at its amino terminus with myristic acid. *Journal of Virology* **61**: 1672-1677.

Petcu, D.J., Aldrich, C.E., Coates, L., Taylor, J.M., Mason, W.S. (1988). Suramin inhibits *in vitro* infection by duck hepatitis B virus, rous sarcoma virus and hepatitis delta virus. *Virology* **167**: 385-392.

Peterson, D.L., Nath, N., Gavilanes, F. (1982). Structure of hepatitis B surface antigen - correlation of subtype with amino acid sequence and location of the carbohydrate moiety. *Journal of Biological Chemistry* **257**: 10414-10420.

Petit, M.-A. and Pillot, J. (1985). HBc and HBe antigenicity and DNA-binding activity of major core protein P22 in hepatitis B virus core particles isolated from the cytoplasm of human liver cells. *Journal of Virology* **53**: 543-551.

Petit, M.-A., Capel, F., Dubanchet, S., Mabit, H. (1992). PreS1-specific binding proteins act as potential receptors for hepatitis B virus in human hepatocytes. *Virology* **187**: 211-222.

Pontisso, P., Petit, M.-A., Bankowski, M.J., Peeples, M.E. (1989). Human liver plasma membranes contain receptors for the hepatitis B virus preS1 region and, *via* polymerized human serum albumin, for the preS2 region. *Journal of Virology* **63**: 1981-1988.

Pourcel, C.H., Louise, A., Gervais, M., Cheneirir, N., Dubois, M.-F., Tiollais, P. (1982). Transcription of hepatitis B surface antigen gene in mouse cells transformed with cloned viral DNA. *Journal of Virology* **42**: 100-105.

Prince, A.M., Ikram, H., Hopp, T.P. (1982). Hepatitis B virus vaccine: identification of HBsAg/*a* and HBsAg/*d* but not HBsAg/*y* subtype antigenic determinants on a synthetic immunogenic peptide. *Proceedings of the National Academy of Sciences of the United States of America* **79**: 579-582.

Proudfoot, N.J. and Brownlee, G.G. (1976). 3′ non-coding region sequences in eukaryotic messenger RNA. *Nature (London)* **263**: 211-214.

Pugh, J.C., Weber, C., Houston, H., Murray, K. (1986). Expression of the X gene of hepatitis B virus. *Journal of Medical Virology* **20**: 229-246.

Pugh, J., Zweidler, A., Summers, J. (1989). Characterization of the major duck hepatitis B virus core particle protein. *Journal of Virology* **63**: 1371-1376.

Purcell, R.H. and Gerin, J.L. (1985). Prospects for second and third generation hepatitis B vaccines. *Hepatology* **5**: 159-163.

Radziwill, G., Zentgraf, H., Schaller, H., Bosch, V. (1988). The duck hepatitis B virus DNA polymerase is tightly associated with the viral core structure and unable to switch to an exogenous template. *Virology* **163**: 123-132.

Radziwill, G., Tucker, W., Schaller, H. (1990). Mutational analysis of the hepatitis B virus P gene product domain structure and RNAseH activity. *Journal of Virology* **64**: 613-620.

Rall, L.B., Standring, D.N., Laub, O., Rutter, W.J. (1983). Transcription of hepatitis B virus by RNA polymerase II. *Molecular and Cellular Biology* **3**: 1766-1773.

Raney, A.K., Milich, D.R., McLachlan, A. (1989). Characterisation of hepatitis B virus major surface antigen gene transcription regulatory elements in differentiated hepatoma cell lines. *Journal of Virology* **63**: 3919-3925.

Raney, A.K., Milich, D.R., Easton, A.J., McLachlan, A. (1990). Differentiation-specific transcriptional regulation of the hepatitis B virus large surface antigen gene in human hepatoma cell lines. *Journal of Virology* **64**: 2360-2368.

Robinson, W.S., Clayton, D.A., Greenman, R.L. (1974). DNA of a human hepatitis B candidate. *Journal of Virology* **14**: 384-391.

Robinson, W.S. and Luttwick, L.I. (1976). The virus of hepatitis type B. *New England Journal of Medicine* **295**: 1168-1175.

Roossinck, M.J., Jameel, S., Loukin, S.H., Siddiqui, A. (1986). Expression of hepatitis B viral core region in mammalian cells. *Molecular and Cellular Biology* **6**: 1393-1400.

Roossinck, M.J. and Siddiqui, A. (1987). *In vivo* phosphorylation and protein analysis of hepatitis B virus core antigen. *Journal of Virology* **61**: 955-961.

Rossner, M.T. (1991). PhD Thesis. University of Edinburgh.

Rossner, M.T. (1992). Hepatitis B virus X gene product - a promiscuous transcriptional transactivator. *Journal of Medical Virology* **36**: 101-117.

Roychoudhury, S. and Shih, C. (1990). *Cis* rescue of a mutated reverse transcriptase gene of human hepatitis B virus by creation of an internal ATG. *Journal of Virology* **64**: 1063-1069.

Ruiz-Opazo, N., Chakraborty, P.R., Shafritz, D.A. (1982). Evidence for supercoiled hepatitis B virus DNA in chimpanzee liver and serum Dane particles: possible implications in persistent HBV infection. *Cell* **29**: 129-138.

Russnak, R. and Ganem, D. (1990). Sequences 5′ to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes and Development* **4**: 764-776.

Russnak, R.H. (1991). Regulation of polyadenylation in hepatitis B viruses - stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive. *Nucleic Acids Research* **19**: 6449-6456.

Sabatini, D.D., Kreibich, G., Morimoto, T., Adesnok, M. (1982). Mechanisms for the incorporation of proteins in membranes and organelles. *Journal of Cell Biology* **92**: 1-22.

Saito, I., Oya, Y., Shimojo, H. (1986). Novel RNA family structure of hepatitis B virus expressed in human cells, using a helper-free adenovirus vector. *Journal of Virology* **58**: 554-560.

Salfeld, J., Pfaff, E., Noah, M., Schaller, H. (1989). Antigenic determinants and functional domains in core antigen and E antigen from hepatitis B virus. *Journal of Virology* **63**: 798-808.

Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463-5467.

Santantonio, T., Jung, M.-C., Schneider, R., Fernholz, D., Milella, M., Monno, L., Pastore, G., Pape, G.R., Will, H. (1992). Hepatitis B virus genomes that cannot synthesise preS2 proteins occur frequently and as dominant virus populations in chronic carriers in Italy. *Virology* **188**: 948-952.

Schek, N., Bartenschlager, R., Kuhn, C., Schaller, H. (1991). Phosphorylation and rapid turnover of hepatitis B virus X protein expressed in HepG2 cells from a recombinant vaccinia virus vector. *Oncogene* **6**: 1735-1744.

Schlicht, H.J., Kuhn, C., Guhr, B., Mattaliano, R.J., Schaller, H. (1987). Biochemical and immunological characterization of the duck hepatitis B virus envelope proteins. *Journal of Virology* **61**: 2280-2285.

Schlicht, H.J. and Schaller, H. (1989). The secretory core protein of human hepatitis B virus is expressed on the cell surface. *Journal of Virology* **63**: 5399-5404.

Schlicht, H.J., Bartenschlager, R., Schaller, H. (1989a). The duck hepatitis B virus core protein contains a highly phosphorylated C terminus that is essential for replication but not for RNA packaging. *Journal of Virology* **63**: 2995-3000.

Schlicht, H.J., Radziwill, G., Schaller, H. (1989b). Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core-polymerase fusion proteins. *Cell* **56**: 85-92.

Schlicht, H.J. and Wasenauer, G. (1991). The quaternary structure, antigenicity and aggregational behaviour of the secretory core protein of human hepatitis B virus are determined by its signal sequence. *Journal of Virology* **65**: 6817-6825.

Schneider, R., Fernholz, D., Wildner, G., Will, H. (1991). Mechanism, kinetics and role of duck hepatitis B virus E antigen expression *in vivo*. *Virology* **182**: 503-512.

Schoedel, F., Moriarty, A.M., Peterson, D.L., Zheng, J., Hughes, J.L., Will, H., Leturcq, D.J., McGee, J.S., Milich, D.R. (1992). The position of heterologous epitopes inserted in hepatitis B virus core particles determines their immunogenicity. *Journal of Virology* **66**: 106-114.

Schoedel, F., Peterson, D., Zheng, J., Jones, J.E., Hughes, J.L., Milich, D.R. (1993). Structure of hepatitis B virus core and E antigen - a single precore amino acid prevents nucleocapsid assembly. *Journal of Biological Chemistry* **268**: 1332-1337.

Schroder, R., Maassen, A., Lippoldt, A., Borner, T., Vonbaehr, R., Dobrowolski, P. (1991). Expression of the core antigen gene of hepatitis B virus (HBV) in *Acetobacter methanolicus* using broad host range vectors. *Applied Microbiology and Biotechnology* **35**: 631-637.

Seeger, C., Ganem, D., Varmus, H. (1986). Biochemical and genetic evidence for the hepatitis B virus replication strategy. *Science* **232**: 477-483.

Seeger, C. and Maragos, J. (1990). Identification and characteristation of the woodchuck hepatitis virus origin of DNA replication. *Journal of Virology* **64**: 16-23.

Seeger, C., Summers, J., Mason, W.S. (1991). Viral DNA Synthesis. *Current Topics in Microbiology and Immunology* **168**: 41-60.

Sells, M.A., Chen, M.-L., Acs, G. (1987). Production of hepatitis B virus particles in HepG2 cells transfected with cloned hepatitis B virus DNA. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 1005-1009.

Seto, E., Zhou, D.X., Peterlin, B.M., Yen, T.S.B. (1989). Trans-activation by the hepatitis B virus X protein shows cell-type specificity. *Virology* **173**: 764-766.

Seto, E., Mitchell, P.J., Yen, T.S.B. (1990). Transactivation by the hepatitis B virus X protein depends on AP-2 and other transcription factors. *Nature (London)* **344**: 72-74.

Shaul, Y., Rutter, W.J., Laub, O. (1985). A human hepatitis B viral enhancer element. *EMBO Journal* **4**: 427-430.

Shikata, T., Karasawa, T., Abe, K., Uzawa, T., Suzuki, H., Oda, T., Imai, M., Mayumi, M., Moritsugu, Y. (1977). Hepatitis B E antigen and infectivity of hepatitis B virus. *Journal of Infectious Diseases* **136**: 571-

Siddiqui, A., Jameel, S., Mapoles, J. (1986). Transcription control elements of hepatitis B surface antigen gene. *Proceedings of the National Academy of Sciences of the United States of America* **83**: 566-570.

Siddiqui, A., Jameel, S., Mapoles, J. (1987). Expression of the hepatitis B virus X gene in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 2513-2517.

Simonsen, C.C. and Levinson, A.D. (1983). Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using Simian Virus 40 - hepatitis B virus chimeric plasmids. *Molecular and Cellular Biology* **3**: 2250-2258.

Sprengel, R., Kaleta, E.F., Will, H. (1988). Isolation and characterisation of a hepatitis B virus endemic in herons. *Journal of Virology* **62**: 3832-3839.

Stahl, S., Mackay, P., Magazin, M., Bruce, S.A., Murray, K. (1982). Hepatitis B virus core antigen: synthesis in *Escherichia coli* and application in diagnosis. *Proceedings of the National Academy of Sciences of the United States of America* **79**: 1606-1610.

Stahl, S.J. and Murray, K. (1989). Immunogenicity of peptide fusions to hepatitis B virus core antigen. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 6283-6287.

Standring, D.N., Rutter, W.J., Varmus, H.E., Ganem, D. (1984). Transcription of the hepatitis B surface antigen gene in cultured murine cells initiates within the presurface region. *Journal of Virology* **50**: 563-571.

Standring, D.N., Ou, J.H., Rutter, W.J. (1986). Assembly of viral particles in *Xenopus* oocytes: Pre-surface antigens regulate secretion of the hepatitis B viral surface envelope particle. *Proceedings of the National Academy of Sciences of the United States of America* **83**: 9338-9342.

Standring, D.N., Ou, J.H., Masiarz, F.R., Rutter, W.J. (1988). A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of E antigens in *Xenopus* oocytes. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 8405-8409.

Stibbe, W. and Gerlich, W. (1983). Structural relationships between minor and major proteins of hepatitis B surface antigen. *Journal of Virology* **46**: 626-628.

Su, H. and Yee, J.-K. (1992). Regulation of hepatitis B virus gene expression by its two enhancers. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 2708-2712.

Su, T.-S., Lai, C.-J., Huang, J.-L., Hin, L.-H., Yauk, Y.-K., Chang, C.-M., Lo, S.-J., Han, S.-H. (1989). Hepatitis B virus transcript produced by RNA splicing. *Journal of Virology* **63**: 4011-4018.

Summers, J., O'Connell, A., Millman, I. (1975). Genome of hepatitis B virus restriction enzyme cleavage and structure of DNA extracted from Dane particles. *Procedings of the National Academy of Sciences of the United States of America* **72**: 4597-4601.

Summers, J., Smolec, J.M., Snyder, R. (1978). A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proceedings of the National Academy of Sciences of the United States of America* **75**: 4533-4537.

Summers, J. and Mason, W.S. (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**: 403-415.

Summers, J., Smith, P.M., Horwich, A.L. (1990). Hepadnavirus envelope proteins regulate covalently closed circular DNA amplification. *Journal of Virology* **64**: 2819-2824.

Sureau, C., Rommet-Lemonne, J.L., Mullins, J.I., Essex, M. (1986). Production of hepatitis B virus by a differentiated cell line after transformation with cloned circular HBV DNA. *Cell* **47**: 37-47.

Sureau, C., Eichberg, J.W., Hubbard, G.B., Romet-Lemonne, J.L., Essex, M. (1988). A molecularly cloned hepatitis B virus produced *in vitro* is infectious in a chimpanzee. *Journal of Virology* **62**: 3064-3067.

Suzuki, M. (1989). SPXX, a frequent sequence motif in gene regulatory proteins. *Journal of Molecular Biology* **207**: 61-84.

Suzuki, T., Masui, N., Kajino, K., Saito, I., Miyamura, T. (1989). Detection and mapping of spliced RNA from a human hepatoma cell line transfected with the hepatitis B virus genome. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 8422-8426.

Suzuki, T., Kajimo, K., Masui, N., Saito, I., Miyamura, T. (1990). Alternative splicing of hepatitis B virus RNAs in HepG2 cells transfected with the viral DNA. *Virology* **179**: 881-885.

Szmuness, W., Stevens, C.E., Hurley, E.J., Zang, E.A., Oleszlo, W.R., Williams, D.C., Sadovsky, R., Morrison, J.M., Kelner, A. (1980). Hepatitis B vaccine. Demonstration of efficacy in a controlled clinical trial in a high risk population in the United States. *New England Journal of Medicine* **303**: 833-841.

Tabor, E. and Gerety, R.J. (1984). Possible role of immune response to hepatitis core antigen in protection against hepatitis B infections. *Lancet* (i): 172.

Tagawa, M., Omata, M., Okuda, K. (1986). Appearance of viral RNA transcripts in the early stage of duck hepatitis B virus infection. *Virology* **152**: 477-182.

Takahashi, K., Akahane, Y., Gotanda, T., Mishiro, T., Imai, M., Miyakawa, Y., Mayumi, M. (1979). Demonstration of hepatitis B E antigen in the core of Dane particles. *Journal of Immunology* **122**: 275-279.

Takahashi, K., Machida, A., Funatsu, G., Nomura, M., Usuda, S., Aoyagi, S., Tachibana, K., Miyamoto, H., Imai, M., Makamura, T., Miyakawa, Y., Mayumi, M. (1983). Immunochemical structure of hepatitis B E antigen in the serum. *Journal of Immunology* **130**: 2903-2907.

Takahashi, K., Kishimoto, S., Ohori, K., Yoshizawa, H., Machida, A., Ohnuma, H., Tsuda, F., Munekata, E., Miyakawa, Y., Mayumi, M. (1991). Molecular heterogeneity of E antigen polypeptides in sera from carriers of hepatitis B virus. *Journal of Immunology* **147**: 3156-3160.

Tay, N., Chan, S.-H., Ren, E.-C. (1992). Identification and cloning of a novel heterogeneous nuclear ribonucleoprotein C-like protein that functions as a transcriptional activator of the hepatitis B virus enhancer II. *Journal of Virology* **66**: 6841-6848.

Terre, S., Petit, M.-A., Brechot, C. (1991). Defective hepatitis B virus particles are generated by packaging and reverse transcription of spliced viral RNAs *in vivo*. *Journal of Virology* **65**: 5539-5543.

Toh, H., Hayashida, H., Miyata, T. (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. *Nature (London)* **305**: 827-829.

Treinin, M. and Laub, O. (1987). Identification of a promoter element located upstream from the hepatitis B virus X gene. *Molecular and Cellular Biology* **7**: 545-548.

Trevisan, A., Realdi, G., Alberti, A., Ongaro, G., Pornaro, E., Meliconi, R. (1982). Core antigen-specific immunoglobulin G bound to the liver cell membrane in chronic hepatitis B. *Gastroenterology* **82**: 218-222.

Tuttleman, J.S., Pourcel, C., Summers, J.W. (1986a) Formation of the pool of covalently closed circular viral DNA in hepadnavirus-infected cells. *Cell* **47**: 451-46 .

Tuttleman, J., Pugh, J., Summers, J. (1986b). *In vitro* experimental infection of primary duck hepatocyte cultures with DHBV. *Journal of Virology* **58**: 17-25.

Twu, J.S. and Schloemer, R.H. (1989). Transcription of human β - interferon gene is inhibited by hepatitis B virus. *Journal of Virology* **63**: 3065-3071.

Twu, J.S., Chu, K., Robinson, W.S. (1989). Hepatitis B virus X gene activates KB-like enhancer sequences in the long terminal repeat of human immunodeficiency virus 1. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 5186-5172.

Ueda, K., Tsurimoto, T., Matsubara, K. (1991). Three envelope proteins of hepatitis B virus: large S, middle S and major S proteins needed for the formation of Dane particles. *Journal of Virology* **65**: 3521-3529.

Urich, R., Borisova, G.P., Siakkou, H., Platzer, C., Ose, V.P., Berzin, J.G., Dreilina, D.E., Pushko, P.M., Tsibinogin, V.V., Pumpen, P.P., Rosenthal, H.A., Gren, E.J. (1991). Surface exposure of a major immunodominant epitope of the gp51 coat protein of Bovine Leukemia Virus on capsids consisting of the hepatitis B virus core antigen. *Molecular Biology* **25**: 301-306.

Uy, A., Bruss, V., Gerlich, W.H., Kochel, H.G., Thomssen, R. (1986). Precore sequence of hepatitis B virus inducing E antigen and membrane-association of the viral core protein. *Virology* **155**: 89-96.

Valenzuela, P., Qurioga, M., Zaldiver, J., Gray, P., Rutter, W.J. (1980). The nucleotide sequence of the hepatitis B viral genome and identification of the major viral genes. In *Animal Virus Genetics: ICN/UCLA Symposium on Molecular and Cellular Biology*, Fields, B.N., Jaenisch, R. and Fox, C.F. eds. (Academic Press, New York), pp.57-70.

Vannice, J.L. and Levinson, A.D. (1988). Properties of human hepatitis B virus enhancer: position effects and cell-type nonspecificity. *Journal of Virology* **62**: 1305-1313.

Wang, G.-H. and Seeger, C. (1992). The reverse transcriptase of hepatitis B virus acts as a protein primer for viral DNA synthesis. *Cell* **71**: 663-670.

Wang, J., Lee, A.S., Ou, J.H. (1991). Proteolytic conversion of hepatitis B virus E antigen precursor to end-product occurs in a postendoplasmic reticulum compartment. *Journal of Virology* **65**: 5080-5083.

Wasenauer, G., Kock, J., Schlicht, H.J. (1992). A cysteine and a hydrophobic sequence in the noncleaved portion of the preC leader peptide determine the biophysical properties of the secretory core protein (HBe protein) of human hepatitis B virus. *Journal of Virology* **66**: 5338-5346.

Weber, C., Bruce, S.A., Peutherer, J.F., Pugh, J.C., Murray, K.M. (1988). Antibodies to the X antigen of hepatitis B virus appear during infection.In *Viral Hepatitis and Liver Disease*, A.J. Zuckerman ed. (Alan R Liss Inc. New York) pp 671-674.

Weimer, T., Salfeld, J., Will, H. (1987). Expression of hepatitis B core gene *in vitro* and *in vivoJournal of Virology* **61**: 3109-3113.

Weiser, B., Ganem, D., Seeger, C., Varmus, H.E. (1983). Closed circular viral DNA and asymmetric heterogeneous factors in livers from animals infected with ground squirrel hepatitis virus. *Journal of Virology* **48**: 1-9.

Whitten, T.M., Quets, A.T., Schloemer, R.H. (1991). Identification of the hepatitis B virus factor that inhibits expression of the beta interferon gene. *Journal of Virology* **65**: 4699-4704.

W.H.O. Scientific Group on Prevention and Control of Hepatocellular Carcinoma (1983). Prevention of primary liver cancer. Report of a meeting of a WHO scientific group. *Lancet* **(i)**: 463-465.

Will, H., Reiser, W., Weimer, T., Pfaff, E., Buscher, M., Sprengel, R., Cattaneo, R., Schaller, H. (1987). Replication strategy of human hepatitis B virus. *Journal of Virology* **61**: 904-911.

Wollersheim, M., Debelka, V., Hofschneider, P.H. (1988). A transcativating function encoded in the hepatitis B virus X gene is conserved in the integrated state. *Oncogene* **3**: 545-552.

Wu, H.-L., Chen, P.-J., Tu, S.-J., Lin, M.-H., Lai, M.-Y., Chen, D.-S. (1991). Characterisation and genetic analysis of alternatively-spliced transcripts of hepatitis B virus in infected human liver tissues and transfected HepG2 cells. *Journal of Virology* **65**: 1680-1686.

Wu, J.Y., Zhou, Z.-Y., Judd, A., Cartwright, C.A,. Robinson, W.S. (1990a). The hepatitis B virus-encoded transcriptional transactivator HBX appears to be a novel protein serine-threonine kinase. *Cell* **63**: 687-695.

Wu, T.-T., Coates, L., Aldrich, C.E., Summers, J., Mason, W.S. (1990b). In hepatocytes infected with duck hepatitis B virus, the template for viral DNA synthesis is amplified by an intracellular pathway. *Virology* **175**: 255-261.

Yaginuma, K., Shirakata, Y., Kobayashi, M., Koike, K. (1987). Hepatitis B virus (HBV) particles are produced in a cell culture system by transient expression of transfected HBV DNA. *Proceedings of the National Academy of Sciences of the United States of America* **84**: 2678-2682.

Yaginuma, K. and Koike, K. (1989). Identification of a promoter region for 3.6 kilobase mRNA of hepatitis B virus and specific cellular binding protein. *Journal of Virology* **63**: 2914-2921.

Yang, S.Q., Walter, M., Standring, D.N. (1992). Hepatitis B virus p25 precore protein accumulates in *Xenopus* oocytes as an untranslocated phosphoprotein with an uncleaved signal peptide. *Journal of Virology* **66**: 37-45.

Yee, J.K. (1989). A liver-specific enhancer in the core promoter region of human hepatitis B virus. *Science* **246**: 658-661.

Yeh, C.-T., Liaw, Y.-F., Ou, J.-H. (1990). The arginine-rich domain of HBV precore and core proteins contains a signal for nuclear transport. *Journal of Virology* **64**: 6141-6147.

Yeh, C.-T. and Ou, J.-H. (1991). Phosphorylation of hepatitis B virus precore and core proteins. *Journal of Virology* **65**: 2327-2331.

Yu, M.S. and Summers, J. (1991). A domain of the hepadnavirus capsid protein is specifically required for DNA maturation and virus assembly. *Journal of Virology* **65**: 2511-2517.

Zahn, P., Hofschneider, P.H., Koshy, R. (1988). The HBV X-ORF encodes a transactivator: a potential factor in viral hepatocarcinogenesis. *Oncogene* **3**: 169-177.

Zelent, A., Sells, M.A., Price, P.M., Mohemad, A., Acs, G., Christman, J.K. (1987). Murine cells carrying integrated tandem genes of hepatitis B virus DNA transcribe RNAs from endogenous promoters on both viral strands and express middle and major viral envelope proteins. *Journal of Virology* **61**: 1108-1115.

Zheng, J., Schodel, F., Peterson, D.L. (1992). The structure of hepadnaviral core antigens: identification of free thiols and determination of the disulphide bonding pattern. *Journal of Biological Chemistry* **267**: 9422-9429.

Zhou, S.L. and Standring, D.N. (1991). Production of hepatitis B virus nucleocapsidlike core particles in *Xenopus* oocytes - assembly occurs mainly in the cytoplasm and does not require the nucleus. *Journal of Virology* **65**: 5457-5464.

Zhou, S., Yang, S.Q., Standring, D.N. (1992). Characterization of hepatitis B virus capsid particle assembly in *Xenopus* oocytes. *Journal of Virology* **66**: 3086-3092.

Appendix: Nucleotide and amino acid sequence of the core gene of HBV subtype *adyw*

The nucleotide sequence of the core gene of HBV subtype *adyw* (Pasek *et al.*, 1979) is shown. Both strands of the DNA are shown, with the sense strand on top. Numbering is by the convention of Pasek *et al.* (1979) which sets number 1 at the start of the core ORF. The amino acid sequence of HBcAg is shown below the DNA sequence.

```
    ATGGACATTGACCCTTATAAAGAATTTGGAGCTACTGTGGAGTTACTCTCGTTTTTGCCT
  1 ---------+---------+---------+---------+---------+---------+ 60
    TACCTGTAACTGGGAATATTTCTTAAACCTCGATGACACCTCAATGAGAGCAAAAACGGA

    MetAspIleAspProTyrLysGluPheGlyAlaThrValGluLeuLeuSerPheLeuPro -

    TCTGACTTCTTTCCTTCCGTACGAGATCTTCTAGATACCGCCGCAGCTCTGTATCGGGAT
 61 ---------+---------+---------+---------+---------+---------+ 120
    AGACTGAAGAAAGGAAGGCATGCTCTAGAAGATCTATGGCGGCGTCGAGACATAGCCCTA

    SerAspPhePheProSerValArgAspLeuLeuAspThrAlaAlaAlaLeuTyrArgAsp -

    GCCTTAGAGTCTCCTGAGCATTGTTCACCTCACCATACTGCACTCAGGCAAGCAATTCTT
121 ---------+---------+---------+---------+---------+---------+ 180
    CGGAATCTCAGAGGACTCGTAACAAGTGGAGTGGTATGACGTGAGTCCGTTCGTTAAGAA

    AlaLeuGluSerProGluHisCysSerProHisHisThrAlaLeuArgGlnAlaIleLeu -

    TGCTGGGGAGACTTAATGACTCTAGCTACCTGGGTGGGTACTAATTTAGAAGATCCAGCA
181 ---------+---------+---------+---------+---------+---------+ 240
    ACGACCCCTCTGAATTACTGAGATCGATGGACCCACCCATGATTAAATCTTCTAGGTCGT

    CysTrpGlyAspLeuMetThrLeuAlaThrTrpValGlyThrAsnLeuGluAspProAla -

    TCTAGGGACCTAGTAGTCAGTTATGTCAACACTAATGTGGGCCTAAAGTTCAGACAATTA
241 ---------+---------+---------+---------+---------+---------+ 300
    AGATCCCTGGATCATCAGTCAATACAGTTGTGATTACACCCGGATTTCAAGTCTGTTAAT

    SerArgAspLeuValValSerTyrValAsnThrAsnValGlyLeuLysPheArgGlnLeu -

    TTGTGGTTTCACATTTCTTGTCTCACTTTTGGAAGAGAAACGGTTCTAGAGTATTTGGTG
301 ---------+---------+---------+---------+---------+---------+ 360
    AACACCAAAGTGTAAAGAACAGAGTGAAAACCTTCTCTTTGCCAAGATCTCATAAACCAC

    LeuTrpPheHisIleSerCysLeuThrPheGlyArgGluThrValLeuGluTyrLeuVal -

    TCTTTTGGAGTGTGGATTCGCACTCCTCCAGCTTATAGACCACCAAATGCCCCTATCCTA
361 ---------+---------+---------+---------+---------+---------+ 420
    AGAAAACCTCACACCTAAGCGTGAGGAGGTCGAATATCTGGTGGTTTACGGGGATAGGAT

    SerPheGlyValTrpIleArgThrProProAlaTyrArgProProAsnAlaProIleLeu -

    TCAACACTTCCGGAGACTACTGTTGTTAGACGACGATGCAGGTCCCCTAGAAGAAGAACT
421 ---------+---------+---------+---------+---------+---------+ 480
    AGTTGTGAAGGCCTCTGATGACAACAATCTGCTGCTACGTCCAGGGGATCTTCTTCTTGA

    SerThrLeuProGluThrThrValValArgArgArgCysArgSerProArgArgArgThr -

    CCCTCGCCTCGCAGACGAAGATCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAA
481 ---------+---------+---------+---------+---------+---------+ 540
    GGGAGCGGAGCGTCTGCTTCTAGAGTTAGCGGCGCAGCGTCTTCTAGAGTTAGAGCCCTT

    ProSerProArgArgArgArgSerGlnSerProArgArgArgArgSerGlnSerArgGlu -

    TCTCAATGT
541 --------- 549
    AGAGTTACA

    SerGlnCys -
```